

**EXPLORING FOREIGN LANGUAGE
ASSESSMENT LITERACY OF PRE-
SERVICE ENGLISH LANGUAGE
TEACHERS**

PhD Dissertation

Aylin SEVİMEL ŞAHİN

Eskişehir 2019

**EXPLORING FOREIGN LANGUAGE ASSESSMENT LITERACY OF PRE-
SERVICE ENGLISH LANGUAGE TEACHERS**

Aylin SEVİMEL ŞAHİN

PhD DISSERTATION

Department of Foreign Language Education

Program in English Language Teaching

Supervisor: Assist. Prof. Dr. Gonca SUBAŞI

Eskişehir

Anadolu University

Graduate School of Educational Sciences

June 2019

JÜRİ VE ENSTİTÜ ONAYI

Aylin SEVİMEL ŞAHİN'in "Exploring Foreign Language Assessment Literacy of Pre-Service English Language Teachers" başlıklı tezi 18.06.2019 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği Programında, Doktora tezi olarak kabul edilmiştir.

Unvanı-Adı Soyadı

İmza

Üye (Tez Danışmanı) : Dr. Öğr. Üyesi Gonca SUBAŞI

Üye : Prof.Dr. Gül DURMUŞOĞLU KÖSE

Üye : Doç.Dr. Işıl KABAKÇI YURDAKUL

Üye : Doç.Dr. İlknur PEKKANLI SAVAŞKAN

Üye : Dr. Öğr. Üyesi İlknur YÜKSEL

Prof.Dr. Handan DEVECİ
Anadolu Üniversitesi
Eğitim Bilimleri Enstitüsü
Müdür Vekili

ABSTRACT

EXPLORING FOREIGN LANGUAGE ASSESSMENT LITERACY OF PRE-SERVICE ENGLISH LANGUAGE TEACHERS

Aylin SEVİMEL ŞAHİN

Department of Foreign Language Education
Program in English Language Teaching
Anadolu University, Graduate School of Educational Sciences, June 2019

Supervisor: Assist. Prof. Dr. Gonca SUBAŞI

Language assessment is an indispensable element of teaching because it has an impact on the quality of education. To better perform language assessment, language teachers must be assessment literate; that is, they must have the understanding of what language assessment is, why it is used, and how it is carried out to foster language teaching/learning. However, it is questionable how effectively language teachers are trained in terms of language assessment in the phase of teacher education. Therefore, the current study is designed to explore foreign language assessment literacy (FLAL) and the assessment training needs of pre-service English language teachers within ‘English Language Testing and Evaluation (ELTE)’ course offered at Anadolu University. Based on a descriptive research, this study collected data from different stakeholders about their opinions, experiences, needs and expectations by means of open-ended questions during 2017-2018 Academic Year: 10 course lecturers from 7 universities, 22 novice EFL teachers as the graduates of Anadolu University, and pre-service ELT teachers as the course-takers (before the course 164 and after the course 146 participants). The documents regarding the ELTE course content of the participant universities were also gathered, and document analysis for them was conducted. As a result of content analysis for open-ended questions, two themes revealed: ‘The FLAL-ELTE Concept’, and ‘ELTE Course-related Issues’. In terms of the concept, though most participants were aware of what it means, they focused only on the notion of testing, summative assessment, the testing purposes of diagnostics and achievement, and the knowledge base of assessment literacy. Considering ELTE course, all the participants found it beneficial and sufficient, and there was a positive change after the course. Yet, most of them underlined that the content was devoted to only language testing and it was too theoretical as also supported by the documents. Moreover, they discussed that the content was deprived of practical side, integrated language testing, interpretation and evaluation skills, the contextual or local needs about language assessment. Thus, there is a need to revise the course content to respond to the assessment training needs of pre-service teachers in order to help them to be more competent with respect to LAL. Accordingly, in the light of the findings, certain implications were shared.

Keywords: English Language Teaching, Language Assessment Literacy, Language Testing, Teacher Education.

ÖZ

İNGİLİZ DİLİ ÖĞRETMEN ADAYLARININ YABANCI DİLDE ÖLÇME-DEĞERLENDİRME OKURYAZARLIĞININ İNCELENMESİ

Aylin SEVİMEL ŞAHİN

Yabancı Diller Eğitimi Anabilim Dalı
İngiliz Dili Eğitimi Programı
Anadolu Üniversitesi, Eğitim Bilimleri Enstitüsü, Haziran 2019

Danışman: Dr. Öğr. Üyesi Gonca SUBAŞI

Dilin ölçülüp değerlendirilmesi, eğitimin kalitesini etkilediğinden öğretimin vazgeçilmez bir unsurudur. Dilin ölçme-değerlendirmesini daha iyi gerçekleştirmek için dil öğretmenlerinin ölçme-değerlendirme okuryazarı olması gerekir; diğer bir deyişle, dil öğretmenleri ölçme-değerlendirmenin ne olduğunu, neden kullanıldığını ve nasıl kullanılırsa öğretimi/öğrenimi geliştirdiğini bilmesi gerekir. Bununla birlikte, öğretmen yetiştirme aşamasında dil öğretmenlerinin dilin ölçme-değerlendirmesi konusunda ne kadar etkili bir şekilde yetiştirildiği tartışmalı bir konudur. Bu nedenle, bu çalışma, Anadolu Üniversitesi'nde verilen 'İngilizce Sınav Hazırlama ve Değerlendirme' dersi kapsamında yabancı dilde ölçme-değerlendirme okuryazarlığını ve İngiliz Dili Eğitimi öğretmen adaylarının bu alandaki eğitim ihtiyaçlarını araştırmak için tasarlanmıştır. Betimsel araştırmaya dayanarak, bu çalışma 2017-2018 akademik yılı boyunca açık uçlu sorularla farklı paydaşlardan onların görüşleri, deneyimleri, ihtiyaçları ve beklentileri hakkında veri toplamıştır: 7 üniversiteden katılan 10 öğretim görevlisi, Anadolu Üniversitesi mezunları olarak mesleğe yeni başlamış 22 İngilizce öğretmeni ve dersi alan İngilizce öğretmen adayları (dersten önce 164 ve dersten sonra 146 katılımcı). Ayrıca, katılımcı üniversitelerin bu dersle ilgili içerikleri hakkında dokümanlar toplanmış ve doküman analizi ile incelenmiştir. Açık uçlu sorular için yapılan içerik analizi sonucunda iki tema ortaya çıkmıştır: 'Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığı Kavramı' ve 'İngilizce Sınav Hazırlama ve Değerlendirme Dersi ile İlgili Konular'. Kavram açısından, çoğu katılımcı kavramın ne olduğunu bilmesine rağmen kavramın sadece test etme, sonuç değerlendirme, tanılayıcı test ve başarı testi, okuryazarlığın bilgi temeli açlarına odaklanmıştır. Ders hakkında ise onu faydalı ve yeterli bulmuşlar ve ders sonrasında olumlu bir değişiklik olmuştur. Buna karşın, doküman analizi sonuçlarının da desteklemesiyle çoğu katılımcı ders içeriğinin sadece dilin ölçülmesi ve test edilmesine ayrıldığının ve dersin çok teorik boyutta kaldığının altını çizmiştir. Ayrıca, ders içeriğinin uygulama yönünden, bütünleşik dil sınavı konusundan, yorumlama ve değerlendirme becerilerinden ve ortama bağlı veya yerel dil ölçme-değerlendirme ihtiyaçlarından yoksun olduğunu tartışmışlardır. Bu yüzden, ders içeriğinin yabancı dilde ölçme-değerlendirme okuryazarlığı bağlamında öğretmen adaylarının daha yetkin olması ve onların bu konudaki eğitim ihtiyaçlarını karşılayacak şekilde gözden geçirilip düzenlenmesi gerekmektedir. Dolayısıyla, bulgular ışığında bazı öneriler paylaşılmıştır.

Anahtar Sözcükler: Dilde Ölçme-Değerlendirme Okuryazarlığı, Dili Test Etme, İngiliz Dili Eğitimi, Öğretmen Eğitimi.

ACKNOWLEDGEMENTS

Writing my PhD dissertation is a long journey which included a number of challenges and great pleasure from its beginning to its completion. This dissertation has been shaped by the efforts and supports of many people to whom I am deeply indebted. Without their encouragement and guidance, this study would not be possible.

First and foremost, I would like to express my very great appreciation to my supervisor Assist. Prof. Dr. Gonca SUBAŞI for her profound knowledge and expertise, continuous assistance, constructive feedback, motivation, inspiration and patience in each phase of writing my research. I should indicate that I feel lucky to have an opportunity to study with such a wonderful academician. I would also like to express my deep gratitude to my dissertation committee members Prof. Dr. Gül DURMUŞOĞLU-KÖSE and Assoc. Prof. Dr. Işıl KABAKÇI-YURDAKUL, who are also my dear teachers and excellent academicians, for their vast knowledge and experience, professional guidance and valuable contributions throughout every stage of my PhD study. Besides, I would like to extend my grateful thanks to my defense committee members Assoc. Prof. Dr. İlknur SAVAŞKAN and Assist. Prof. Dr. İlknur YÜKSEL for their insightful comments, useful suggestions and widening my research from various perspectives.

I also wish to acknowledge that I received a great deal of help and contribution from various people in composing my dissertation. I would like to offer my grateful thanks to Inst. Neslihan AYDEMİR for her expert opinion, feedback and help during data collection and analysis. Moreover, I would like to express my sincere gratitude to Assoc. Prof. Dr. Ali MERCİ, Assoc. Prof. Dr. Özgür YILDIRIM, Inst. Ela AKGÜN-ÖZBEK, Inst. Yücel ÖZ and the registrar's office staff at Faculty of Education at Anadolu University for their great assistance especially during data collection. Furthermore, I very much appreciate the great help and valuable contribution of all the participants (the course lecturers from different universities across Turkey, the alumni of Anadolu University as novice EFL teachers, and pre-service ELT teachers of Anadolu University) who deserve the deepest thanks for accepting to be a part of my research by sharing their opinions and experiences to enrich my study. I am very grateful to all for their cooperation and contribution.

The last but not the least, I owe special thanks to my beloved life companion, my husband Hamdullah ŞAHİN for his endless love, respect, great patience, understanding,

kindness, encouragement and support all the time. I am sure that he is always and forever with me. Thank you, my love, for everything.

Aylin SEVİMEL-ŞAHİN

Eskişehir 2019

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with “scientific plagiarism detection program” used by Anadolu University, and that “it does not have any plagiarism” whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

Aylin SEVİMEL-ŞAHİN

THE LIST OF CONTENTS

	<u>Page</u>
COVER PAGE	i
FINAL APPROVAL FOR DISSERTATION	ii
ABSTRACT	iii
ÖZ	iv
ACKNOWLEDGEMENTS	v
STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES	vii
THE LIST OF CONTENTS	viii
THE LIST OF TABLES	xi
THE LIST OF FIGURES	xii
THE LIST OF ABBREVIATIONS	xiv

CHAPTER 1

1. INTRODUCTION	1
1.1. Background to the Study	1
1.2. Statement of the Problem	5
1.3. Purpose of the Study	7
1.4. Significance of the Study	8

CHAPTER 2

2. REVIEW OF LITERATURE	11
2.1. A Brief History and Approaches to Language Testing and Assessment...11	11
2.2. Basic Terms and Concepts: Testing, Assessment and Evaluation	13
2.2.1. Testing	14
2.2.1.1. <i>Definition of testing and measurement</i>	14
2.2.1.2. <i>Purposes and types of tests and testing</i>	15
2.2.1.3. <i>Principles or criteria of testing</i>	18
2.2.2. Assessment	23
2.2.2.1. <i>Definition of assessment</i>	23
2.2.2.2. <i>Types of assessment</i>	24

	<u>Page</u>
2.2.3. Evaluation	28
2.2.4. Other associated terms and concepts	29
2.2.5. Foreign language testing, assessment and evaluation	32
2.2.5.1. <i>Assessing grammar</i>	35
2.2.5.2. <i>Assessing vocabulary</i>	36
2.2.5.3. <i>Assessing reading</i>	37
2.2.5.4. <i>Assessing listening</i>	37
2.2.5.5. <i>Assessing speaking</i>	38
2.2.5.6. <i>Assessing writing</i>	39
2.2.5.7. <i>Assessing young language learners</i>	40
2.2.6. Importance of testing, assessment and evaluation in education ...	40
2.3. Language Assessment Literacy (LAL)	42
2.3.1. The concept of assessment literacy (AL)	42
2.3.2. The concept of language assessment literacy (LAL)	44
2.3.2.1. <i>Definitions of LAL</i>	45
2.3.2.2. <i>Dimensions or components of LAL</i>	48
2.3.2.3. <i>Characteristics of LAL</i>	51
2.3.3. Importance of LAL in foreign language education	54
2.3.4. Review of relevant research studies on LAL	56
2.3.4.1. <i>Research studies conducted in in-service EFL contexts</i>	56
2.3.4.2. <i>Research studies conducted in pre-service EFL contexts</i>	80
2.4. Training of Language Assessment Literacy (LAL)	85
2.4.1. Importance of LAL training	85
2.4.2. LAL training in EFL/ESL contexts across the world	87
2.4.3. LAL training in Turkey	94
2.4.3.1. <i>Content of LAL training in Turkey</i>	95
2.4.3.2. <i>Research studies on LAL training in Turkey</i>	97

CHAPTER 3

3. METHODOLOGY	100
3.1. Type of Research	100
3.2. Research Context	101

	<u>Page</u>
3.3. Participants	102
3.4. Data Collection Procedure	105
3.5. Data Analysis	107

CHAPTER 4

4. FINDINGS AND DISCUSSION	111
4.1. Findings from the Participant Stakeholders and the Documents	111
4.1.1. Findings from course lecturers	111
4.1.2. Findings from novice EFL teachers	121
4.1.3. Findings from pre-service ELT teachers	132
4.1.4. Findings from the documents about ELTE course content	146
4.2. Summary of the Findings	148
4.2.1. The explanation of the first research question	152
4.2.2. The explanation of the second research question	153
4.3. Discussion	155
4.3.1. FLAL-ELTE concept	155
4.3.2. ELTE course-related issues	159
4.3.2.1. <i>The content and practices of ELTE course</i>	159
4.3.2.2. <i>The sufficiency and shortcomings of ELTE course</i>	162
4.3.3. LAL training needs	168

CHAPTER 5

5. CONCLUSION	173
5.1. Summary of the Study	173
5.2. Implications	175
5.3. Suggestions for Further Research	180

REFERENCES	181
------------------	-----

APPENDICES

CURRICULUM VITAE

THE LIST OF TABLES

	<u>Page</u>
Table 2.1. Research studies conducted in in-service EFL contexts about LAL within foreign contexts	66
Table 2.2. Research studies conducted in in-service EFL contexts about LAL in Turkish context	77
Table 2.3. Research studies conducted in pre-service EFL contexts about LAL in foreign contexts	81
Table 2.4. Research studies conducted in pre-service EFL contexts about LAL in Turkish context	84
Table 2.5. Research studies conducted in EFL/ESL contexts across the world about LAL training	92
Table 2.6. Research studies conducted in Turkish context about LAL training	98
Table 3.1. The distribution of the participants (numbers).....	103
Table 3.2. The profile of the course lecturers	104
Table 3.3. The profile of the novice EFL teachers (the alumni) (numbers)	104
Table 3.4. The perceived levels of ELA by the pre-service teachers after the course (numbers)	105
Table 3.5. Acronyms used for the owners of quotes	110
Table 4.1. The main findings of the document analysis	147
Table 4.2. The comparison and contrast of the ELTE course content	171

THE LIST OF FIGURES

	<u>Page</u>
Figure 2.1. The relationship among the basic terms	14
Figure 2.2. Language assessment literacy: An expanded definition (by courtesy of Fulcher, 2012, p. 126)	49
Figure 2.3. Levels of AL/LAL differentiated according to stakeholder constituency (by courtesy of Taylor, 2013, p. 409)	50
Figure 2.4. Differential AL/LAL profiles for four stakeholders (by courtesy of Taylor, 2013, p. 410)	51
Figure 2.5. The suggested course content for ELTE by CoHE	96
Figure 3.1. The sample screenshot for the categorical themes	108
Figure 4.1. The thematic display of the findings from the course lecturers	112
Figure 4.2. The thematic display of the theme ‘FLAL-ELTE Concept’ for course lecturers	113
Figure 4.3. The thematic display of the subtheme ‘Experiences about ELTE Course’ for course lecturers	116
Figure 4.4. The thematic display of the subtheme ‘General Opinions about ELTE Course’ for course lecturers	119
Figure 4.5. The thematic display of the findings from the novice EFL teachers	122
Figure 4.6. The thematic display of the theme ‘FLAL-ELTE Concept’ for novice EFL teachers	123
Figure 4.7. The thematic display of the subtheme ‘ELTE Course-related Experiences (Previous Experiences)’ for novice EFL teachers	126
Figure 4.8. The thematic display of the subtheme ‘Practices of Using ELTE in Career Life (Current Experiences)’ for novice EFL teachers	130
Figure 4.9. The thematic display of the findings from pre-service ELT teachers (before the course)	133
Figure 4.10. The thematic display of the theme ‘FLAL-ELTE Concept’ for pre-service ELT teachers (before the course)	134

	<u>Page</u>
Figure 4.11. The thematic display of the theme ‘ELTE Course-related Issues’ for pre-service ELT teachers (before the course)	135
Figure 4.12. The thematic display of the findings from pre-service ELT teachers (after the course)	139
Figure 4.13. The thematic display of the theme ‘FLAL-ELTE Concept’ for pre-service ELT teachers (after the course)	140
Figure 4.14. The thematic display of the theme ‘ELTE Course-related Issues’ for pre-service ELT teachers (after the course)	142
Figure 4.15. The word cloud of common keywords found in all the findings	148
Figure 4.16. The thematic display of the findings for all the participants	149

THE LIST OF ABBREVIATIONS

BA	: Bachelor of Arts Degree
CEFR	: Common European Framework Reference for Language
CL	: Course Lecturer
CoHE	: Council of Higher Education
DA	: Dynamic Assessment
EFL	: English as a Foreign Language
ELA	: English Language Assessment
ELT	: English Language Teaching
ELTE	: English Language Testing and Evaluation
ESP	: English for Specific Purposes
FLAL	: Foreign Language Assessment Literacy
FLAL-Eng	: Foreign Language Assessment Literacy in English
LAL	: Language Assessment Literacy
MA	: Master of Arts
MoNE	: Ministry of National Education
N/A	: Not Applicable
OMB-Tr	: ‘Measurement and Evaluation in Education’ - Turkish Course
Ss	: Students
T/Ts	: Teacher/Teachers
Tr	: Turkish

CHAPTER 1

1. INTRODUCTION

1.1. Background to the Study

Teaching and assessment are considered to be two interrelated components in an educational context. In other words, assessment is acknowledged to be an essential element of teaching because it gives feedback about the quality of teaching. Therefore, assessment is helpful to reveal whether teaching is useful or not, whether students learn the targeted behaviors/knowledge/skills, and the extent the objectives of a course is achieved, and so on (Herrera & Macias, 2015; Rogier, 2014). Correspondingly, teachers make decisions about content, materials, alternatives, and the like in order for better and effective teaching practices and instruction (Rea-Dickins, 2004). Thus, since assessment informs teaching, it guides, regulates and reinforces teaching. As a result, students are motivated to learn, and both teachers' and students' performances are improved. Then, it can be inferred that assessment is the bridge between curriculum and teaching.

Teachers are assumed to be the most important factor that impacts teaching/learning. Their responsibilities, knowledge and implementations are not limited to only teaching. One of the teaching competencies teachers possess is the understanding and practice of classroom assessment. Therefore, it can be asserted that teachers have dual roles: being an instructor, and at the same time, an assessor (Inbar-Lourie, 2013b; Rea-Dickins, 2004; Wach, 2012). In other words, while teachers carry out their role of transferring knowledge to their students in their area of specialization, i.e. English language teaching, they also observe and track changes in learning progress, and monitor learning development. However, as for foreign language teachers, it can be assumed that they have three essential responsibilities: being proficient in that foreign language (language proficiency), having the knowledge and expertise in teaching that language (language instruction), and being able to monitor and assess students' language development (language assessment). Among these responsibilities, language assessment is prominent because it shapes teachers' decision-making processes about their language instruction in line with language learning progress (Inbar-Lourie, 2013a; Oz & Atay, 2017). That is to say, as a result of teachers' assessment process, they make decisions about how to evaluate learning outcomes in order to find out what needs to be fulfilled, and how to enhance the missing parts in the learning progress.

Regarding language assessment, every teacher should know what to assess, how to assess, how to manage assessment procedures, how to evaluate and plan their instruction according to the results, how to give feedback, what the rationale and purposes of assessment are, and how to support, affect and regulate learning via assessment in a foreign/second language learning/teaching environment (Hidri, 2018; Scarino, 2013; Wach, 2012). Hence, it can be deduced that teachers must have the necessary skills or the ability to assess as well as the relevant knowledge about assessment procedures to maintain and support the teaching/learning process (DeLuca & Klinger, 2010; Howerton, 2016). That is, teachers are supposed to be familiar with goals, formats, instruments, development and administration procedures, and analyses of assessment to make sense of its results to draw inferences about learning progress and development, the further needs of learners, and the efficiency of the instruction (Inbar-Lourie, 2008). In order to have such characteristics and carry out an effective assessment process, foreign language teachers must be assessment literate.

Language assessment literacy (LAL) is acknowledged to be a certain qualification of teachers' professional competence. Basically, the concept of LAL can be defined as the knowledge, skills, and principles about test construction, development, administration, use, interpretation, evaluation, and analysis in addition to the functions and visions of assessment in a broader sense (Lam, 2014; Pill & Harding, 2013). Since foreign language teachers are required to have an additional competency, which is language assessment, together with language proficiency and content pedagogical knowledge of language teaching, LAL is distinctive from general teacher assessment literacy. Put it differently, the concept of LAL is separated from assessment literacy due to its component of language specific competencies; that is, the target group and the context are different.

To illustrate the difference better, assessment literacy and language assessment literacy can be compared by means of the starting point. From the view of general educational perspective, teacher assessment literacy is defined as the understanding of the concepts of assessment, the ability to use assessment methods and instruments, the ability to interpret the results as to influence teaching decisions, and as a consequence, to give feedback about the quality of teaching and learning (DeLuca & Klinger, 2010; Howerton, 2016; Khadijeh & Amir, 2015; Popham, 2011). In line with this description, certain researchers expanded the definition of assessment literacy to LAL while developing the

frameworks or conceptualizations directly related to and representing foreign or second language contexts. For instance, Davies (2008) discussed that LAL comprised of three aspects: *skills* (the training in assessment methodology such as item writing and statistics), *knowledge* (the background in language description, context, and measurement), and *principles* (use of proper tests, fairness, ethics, and impact). Moreover, Inbar-Lourie (2008) defined LAL as “the understanding the ‘*what*’ and performing the ‘*how*’ necessitates appreciation of the background and reasoning behind the actions taken, that is, the ‘*why*’” (p. 390). Thus, in essence, LAL comprises three main components: the why (reasoning or rationale; theoretical framework), the what (the description of the trait to be assessed; the content), and the how (the assessment process; tools, analysis, reporting, ethics, etc.), out of which the component ‘what’ is emphasized because it is considered as the major variable that assesses the content/the target language (Inbar-Lourie, 2013b). For this reason, according to Inbar-Lourie (2013a; 2013b), LAL is multi-layered in that it requires the knowledge of general education assessment accompanied by language specific concerns and practices such as the familiarity with current approaches in foreign language education. In addition, Fulcher (2012) provided an expanded definition of LAL as a result of his research project on the development of survey instrument of LAL:

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

In view of Fulcher’s (2012) comprehensive definition, LAL basically consists of three dimensions: *practices* (knowledge, skills and abilities), *principles* (processes, principles and concepts), and *contexts* (historical, social, political, and philosophical frameworks: origins, reasons and impacts). It can be concluded that more or less, all the definitions of LAL share basic elements such as skills, abilities, principles, knowledge, and understanding. Hence, language assessment literate teachers are said to have some common characteristics: the understanding of the assessment principles, the knowledge of how to assess, and interpret student learning, the skills of developing, administering, and analyzing the tests, the ability to integrate the outcomes of assessment in instruction,

the ability to use assessment results and link them to approaches, methods and techniques of teaching, and the familiarity with assessment concepts or tools within a particular context, i.e., foreign language learning (Berry & O'Sullivan, 2016; Huang & He, 2016; Khajideh & Amir, 2015; Rogier, 2014; Taylor, 2009). As the requirements of effective teaching are considered, foreign language teachers, then, need to possess LAL. It is important because it is helpful to guide instruction, to give feedback about teaching/learning, to organize the instruction to improve learning, to select appropriate instruments suitable to learning outcomes, and to develop professionally (Buyukkarci, 2016; Khajideh & Amir, 2015).

Because of such importance, there have been recent attempts to investigate LAL to describe its role, the essential components and the practical concerns. In fact, the investigation of LAL has emerged from the growing need for teachers to be competent educators in the field of testing, and also in the effective teaching instruction (Buyukkarci, 2016). Also, there has been a shift in language teaching/learning approaches towards more sociocultural theories of learning, especially with Vygotsky's notions of zone of proximal development (ZPD) and dynamic assessment, in terms of classroom-based assessments (Fulcher, 2012; Inbar-Lourie, 2008; Scarino, 2013). That is, such learning theories have caused primary changes regarding the conceptions of language assessment: A shift from traditional testing notion (summative) to the present conception of assessment (formative) (Hidri, 2018; O'Loughlin, 2006; Tsagari & Vogt, 2017). Besides, there was much emphasis on how to teach languages, and the processes of language learning in the past whereas language assessment was ignored (Viengsang, 2016). Consequently, with the mentioned shift in the teaching/learning theories as well as the changing needs and expectations of current language teaching, the competencies and practices of language assessment have gained a crucial role to enhance learning (Inbar-Lourie, 2008; Scarino, 2013; Scarino, 2017). That is, the concept of language assessment literacy is now more important because of its raised awareness in the field of foreign language education.

In accordance with this increased awareness, to promote LAL of pre-service teachers has begun through training programs. It is because 'teachers are not born testers' (Jin, 2010, p. 556), and they need some kind of training in this respect. Therefore, the development of LAL has gained much importance in teacher education programs. It is also because so far, the qualification of assessment has been ignored, and overshadowed

by methodology education. In order to produce capable, motivated, confident, effective, and competent teachers for their future careers, to provide assessment education with teacher candidates has become more crucial. To conclude, in order to equip teachers with LAL and to improve this competence, teacher education programs have the paramount responsibility due to their training role.

1.2. Statement of the Problem

Language assessment has become a perennial topic in foreign language education after its importance has been noticed and regarded as essential as other teaching competencies. Being aware of and knowledgeable about assessment procedures, testing concepts, and able to use various assessment tools according to different contexts require teachers to be assessment literate in order to monitor and motivate learner progress as well as to feed teaching performance (DeLuca & Klinger, 2010; Khadijeh & Amir, 2015; Rogier, 2014). In other words, without LAL, teachers may not be helpful to their students to reach higher levels of achievement, and hence, may cause a decrease in the quality of education (Howerton, 2016; Huang & He, 2016). Besides, how teachers perceive assessment and how they are good at this can affect the way they teach, and thereby, the way they assess (Jannati, 2015). Thus, to develop LAL of teacher candidates is believed to be a necessary component of foreign language teacher training program (DeLuca & Klinger, 2010; Herrera & Macias, 2015). The training of LAL, then, can encompass ‘an appropriate balance of technical know-how, practical skills, theoretical knowledge, and understanding of principles, but all firmly contextualized within a sound understanding of the role and function of assessment within education and society’ (Taylor, 2009, p. 27).

However, though the necessary skills and knowledge about language assessment have been given in the teacher preparation programs, it is questionable how well they trained their undergraduates as future teachers in this aspect (e.g. Berger, 2012; Chen, 2005; Howerton, 2016; Hudaya, 2017; Malone, 2008; Lam, 2014; Rogier, 2014; Vogt & Tsagari, 2014). This issue has been more debatable than ever because most of the research studies showed that training programs were not successful in supporting LAL of pre-service teachers (Howerton, 2016; Rogier, 2014), and many in-service teachers graduated with poor ability to prepare and practice testing tools, and to assess learner language development; that is, with insufficient levels of LAL (Howerton, 2016; Vogt & Tsagari, 2014). In the studies, the need for extra training was also stated by a number of in-service

teachers because they felt unprepared to undertake assessment practices (Malone, 2008; Rogier, 2014). Besides, another problem with assessment training was found as being unable to interpret the outcomes of assessment so as to improve teaching/learning (Lam, 2014). Therefore, the demand to understand the assessment needs of pre-service language teachers has been highlighted.

This issue is not so different in the Turkish context. In Turkey, most of the graduates from ELT training programs are appointed as EFL teachers by Ministry of National Education (MoNE) to teach English at state schools. According to MoNE-ELT Competencies (2008), English language teachers in Turkey are expected to be qualified in assessment and monitoring, which include the determination of measurement and evaluation aims, to use the instruments and methods of measurement and evaluation, the ability to interpret the results of assessment and to give feedback to determine the language development, and the ability to use the results of assessment in teaching practices. Namely, they are expected to be trained in classroom-assessment; that is, English language teachers prepare and administer assessments, and checking learners' progress on their own in their classrooms. Furthermore, they are expected to be trained in assessing foreign language skills and components of students, which implies that foreign language assessment has its own contextual factors and background for assessment and evaluation. Therefore, as Buyukkarci (2016) has asserted that the training of foreign language teachers is different compared to other fields of education. In this sense, though very few but significant research studies were conducted within the scope of language assessment in Turkey, they reached the conclusion that teacher training programs remain inadequate to meet the assessment needs of their prospective teachers (e.g. Koksall, 2004; Hatipoglu, 2010; Hatipoglu, 2015; Mede & Atay, 2017). Those studies were about testing materials and practices, and the evaluation of language courses given at one teacher education program, but they did not underline how to enhance LAL in teacher education programs.

Considering the evolving or changing needs of pre-service language teachers due to the challenges of recent assessment developments in education, teacher education programs must evaluate, and renew their assessment courses to meet such needs, and to prepare their teachers for their professional teaching career better (Hatipoglu, 2010, 2015; Jeong, 2013) because despite sharing the prescribed curriculum, language assessment courses can be delivered differently, even these courses with the same name may follow

different syllabi (Berger, 2012; Hatipoglu, 2010; Inbar-Lourie, 2008). When English language teacher education programs are taken into consideration in Turkey, there is only one course about testing and evaluation called ‘English Language Testing and Evaluation (ELTE)’. Though its syllabus is set by Council of Higher Education (CoHE), course lecturers follow their own syllabus that is designed as a result of making adaptations. ELTE course is delivered in the last semester of the program and its content is mainly about classroom-based assessment, but its sufficiency is open to discussion (Koksal, 2004; Hatipoglu, 2010, 2015; Sariyildiz, 2018). Therefore, the first step might be to detect the gaps in terms of language assessment literacy (Malone, 2008). More specifically, to determine the assessment needs of pre-service language teachers, to investigate whether the course of testing and evaluation is adequate to enhance LAL, and as a result, to regulate the present syllabus can be the guide for how to provide such training (Fulcher, 2012; Hatipoglu, 2010; Huang & He, 2016).

In addition to such problems and needs, there has been a lack of or a limited number of works in the field of assessment literacy, especially language assessment literacy in relation to teacher education (Hatipoglu, 2010; Huang & He, 2016; Inbar-Lourie, 2008; Montee, Bach, Donovan & Thompson, 2013; Zolfaghari & Ahmadi, 2016). That is, most researchers have studied in-service EFL teachers’ LAL; however, the research into pre-service teachers’ LAL has been only recently started. So, this kind of research, as Fulcher (2012) pointed out, has just begun to develop, much research is required to enlighten the concept of LAL, specifically, about its training side.

All in all, it is clear that there is a need to understand the language assessment needs of pre-service English language teachers and to evaluate the course-based training of LAL in teacher education programs in order to describe the present status of the LAL training, its content, its implementations, and problems. To investigate these issues may be useful to revise the current syllabus to meet the needs, to prevent the future problems of assessment practices, to promote LAL, and to train pre-service English language teacher as being capable of assessment procedures.

1.3. Purpose of the Study

The present study primarily aims to find out whether the course of English language assessment is useful to develop LAL of pre-service English language teachers at Anadolu University through the descriptions, perceptions, and opinions of course-takers, course

lecturers and graduates as novice teachers. It also intends to reveal what they know about LAL, whether they feel prepared to assess, to elicit their assessment training needs, to investigate the efficiency of the syllabus in terms of providing a good model of assessment training, and whether it leads to a change with respect to LAL, and the problems for further training and assistance. To this end, two research questions were addressed:

1. What are the assessment training needs of pre-service English language teachers?
2. How does English language testing and evaluation course affect language assessment literacy of pre-service English language teachers?

1.4. Significance of the Study

The important role of language assessment literacy in education has already been acknowledged. It is because the knowledge of assessment affects the quality of instruction; so, teachers need to be assessment literate as their professional competence to respond to their contextual teaching/learning needs and to improve both teachers' and students' performance (Howerton, 2016; Huang & He, 2016; Khadijeh & Amir, 2015). However, as mentioned previously, there has been little research into this concept in the literature (Fulcher, 2012; Inbar-Lourie, 2013a). Therefore, to study on LAL may contribute to the field of such research, and it might reveal new perspectives with regard to the research context (e.g. Turkey, and English Language Teaching Department).

In addition, it may help to evaluate its training aspect in terms of how it is effective to equip its course-takers with necessary knowledge and skills before starting to teach. This is important because of the fact that there has been a greater demand in language assessment courses in teacher education programs recently (Berger, 2012; Djoub, 2017; Jeong, 2013). It is also because of the recent shift in the notion of testing and assessment (from summative to formative), which makes language assessment courses need to be redesigned.

Similarly, the findings may be useful to reveal the ways to support initial pre-service training on assessment (Dunkley, 2015; Huang & He, 2016; Jin, 2010; Malone, 2008). Therefore, the investigation of the effect of the assessment course on LAL may give some guidelines for future steps. In other words, to find out whether it promotes or inhibits the development of LAL can show new ways to deal with assessment courses.

Furthermore, language teacher candidates need good and effective practices of language assessment early in learning so that they construct their assessment knowledge and skills for their future career (Giraldo, 2018; Herrera & Macias, 2015). Therefore, this study also tries to find out whether such practices are modelled by language assessment course lecturers, and accordingly, it might suggest the ways to improve experiences of course-takers.

Moreover, the discovery of the assessment needs may guide the instruction, and coursework to develop LAL. Since teachers may have different background, abilities and perceptions (Zolfaghari & Ahmadi, 2016), and there are some contextual factors such as educational policies, sociocultural values and local challenges which have an impact on learners and teachers' perceptions and beliefs about assessment and how they shape language assessment knowledge in their minds (Yan, Zhang & Fan, 2018), their training needs can naturally differ.

Likewise, the outcomes of the present study are also supposed to influence the pedagogic decision-making process (Jannati, 2015), and to have a role in the construction of new materials for the course (Fulcher, 2012) because it collects data from different stakeholders as well as documents. Hence, it may reveal what the stakeholders think about the course-based training.

Besides, it might provide an evaluation of assessment courses given in ELT departments of Turkey; such evaluation may be helpful to compare and contrast the contents, and to add or remove some components in the syllabus in correspondence with the assessment needs of pre-service English language teachers. As Hatipoglu (2015) has pointed out, in a way, the findings may show what happens in language assessment course as a part of the teacher preparation program.

When the current study is considered on a conceptual basis, it is believed to shed some light on LAL of pre-service English language teachers in Turkey where English is learned as a foreign language, and where examination and assessment have a great impact on both teachers and students due to its being an exam-oriented country. Thus, there may occur a new conceptualization of LAL or might lead to a revision of the existing frameworks about LAL. Specifically, to study on English language teachers' needs in language assessment may also contribute to the field regarding the comprehension of LAL within the scope of English language teacher education. It is because foreign language teachers' assessment literacy includes the competencies specific to foreign

language teaching/learning and their assessment ways (Buyukkarci, 2016). Therefore, it may yield a better interpretation of the concept of LAL, and hence, a better understanding of the teacher education programs with respect to language assessment course.

To sum up, the present study is hoped to contribute to the field of language assessment research in many ways. Since the pre-service education is important for future professional life of teaching (DeLuca & Klinger, 2010), to find out the needs of pre-service English language teachers is important to guide the instruction. In order to develop their LAL, this study is believed to provide some insights into how to enhance this literacy, and how to design and organize the language assessment education at universities. The last but not the least, the major contribution of the study might be its uncovering the relationship between language assessment training and the development of LAL within the teacher education programs.

CHAPTER 2

2. REVIEW OF LITERATURE

2.1. A Brief History and Approaches to Language Testing and Assessment

Language testing and assessment are integral part of teaching/learning process because they guide and regulate teaching practices by means of providing feedback about the process. Accordingly, teachers plan and carry out their instructional practices in order to increase students' achievement and promote language learning.

However, language testing and assessment were overshadowed by teaching methodologies in the past because tests were considered something separated from teaching (Viengsang, 2016). Its importance has been realized with the emphasis on the testing-teaching relationship in which tests can be used to reinforce and motivate learning as well as to determine students' performances (Heaton, 2011). Therefore, tests and assessments have been acknowledged to be the ways to improve language learning, and the impact of tests and assessments on students, teachers and society have also been recognized.

In fact, when considered historically, it is seen that testing and assessment have been shaped in line with the changing approaches, paradigms, notions or methodologies to language teaching/learning. It is because the notion of assessment is dynamic, which has been adjusted to the changes and needs in society, culture, economy, policy and epistemology throughout the history of language teaching (Berger, 2012). Seeing that language teaching/learning has been framed within behaviorism (1950s) earlier, and then, has been transformed into more communicative approaches (1970s-after), language testing and assessment have also changed correspondingly from more traditional ways to more communicative and learning-centered forms (Brown, 2004). In this issue, for example, Berger (2012) and Uyaniker (2017) have argued that according to Spolsky, there are three phases of language testing: Pre-scientific (traditional) (before 1960s); psychometric-structuralist (modern) (1960s); psycholinguistic-sociolinguistic (integrative/postmodern) (1970s). Likewise, as Jin (2010) has put forward that Spolsky's division of language testing periods reflects the shift from more traditional ways measuring only language proficiency as the product to more formative ways to evaluate the performance and progress of learners in terms of language development. In other words, the shift from such notions that support assessment of learning to more

communicative and learning-oriented concepts that advocate assessment for learning in the last decades (1990s-after) have led different testing and assessment approaches to emerge such as alternative assessment, dynamic assessment and task-based ESP (English for Specific Purposes) testing (Brown, 2004; Campbell & Duncan, 2007; Uyaniker, 2017).

In addition to such argumentations, some researchers have chronologically classified approaches to language testing into four categories: While Brown (2004) has discussed discrete point vs integrative testing, communicative testing and performance-based testing, Heaton (2011) has argued essay-translation, structuralist, integrative and communicative testing.

To begin with the classification of Heaton (2011), *essay-translation approach* of language testing is about the pre-scientific era of testing as mentioned previously. The decisions about students' linguistic knowledge are only bound to teachers' own judgments; so, it is subjective. Some of the sample tests in this approach include writing and translating essays and analyzing grammatical errors in a text. However, *structuralist approach* is dealt with the mastery of separate target language elements such as grammar, vocabulary and phonology. It is a part of the psychometric-structuralist era of testing, which highlights language learning is acquiring and setting correct forms of language habits. In this approach, the separation of language units is essential, and objective testing which is provided via reliable and valid psychometric tests has a paramount importance. Also, test-takers mostly study on structural linguistics and contrastive analysis. On the contrary to the separation of language elements as in structuralist approach, *integrative approach* of language testing supports the idea of the unison of language elements in a context. According to this approach, context is so important that meaning and discourse are assessed through such tasks as cloze tests, oral interviews, composition writing and dictation. Thus, more than one language skill requires to be used at the same time to fulfill the testing tasks, which shows whether learners are able to use the target language correctly. The final approach is *communicative approach* which is fundamentally interested in how language is used by learners in communication. Rather than language usage (correct use of linguistic aspects), language use (meaningful use of language for several purposes like in real life) is underlined in this approach. Communicative testing, then, include communicative, authentic and real-life tasks which represent the culture of the target language. Moreover, in this approach, the division of language skills is essential

because it is assumed that separate points taken for each language skill reflect overall performance of language learners. Through this approach, language bands have also arisen. They are the qualitative descriptions of any language skill which serve as the making-judgments-evaluations according to the performance of language learners.

As for Brown's (2004) classification, *discrete point language testing* is based on the idea of disconnected parts of language components and such tests include all the units of language components, but in an unattached way. On the other hand, *integrative language testing* concentrates on the combination of certain language skills and areas, but mostly linguistic aspects while test-takers need to consider, for instance, grammar and vocabulary in a context at the same time such as cloze tests and dictation. Regarding *communicative language testing*, as the name reveals, it involves authentic or real-life language tasks to assess language learners' overall performance in terms of both linguistically and meaningfully; and thus, to assess whether learners are able to use the target language. Thus, it focuses on language skills such as reading, writing, listening and speaking whereas grammar, vocabulary and pronunciation are embedded in such skills. His last category is about *performance-based testing* which relies mainly on interactive tasks excluding pen-and-paper tests to assess oral and written production, group performance and the like, and it assesses whether learners are able to perform real-world tasks; that is, to elicit communicative performance at the expected level. For performance-based testing, Coombe (2018) has also argued that not the competence but the performance is tested and such features as intonation, gestures, mimes, ideas and organization are taken into account while evaluating the language performance.

Besides, Brown (2004) has further discussed the current applications in language testing. He has exemplified the notion of multiple intelligences, alternative assessment and computer-based testing, all of which have been influenced by the recent perspectives and implementations in language teaching/learning.

All in all, it can be concluded that language testing and assessment have changed in congruence with the shifts in language teaching and learning approaches and methodologies in the history.

2.2. Basic Terms and Concepts: Testing, Assessment and Evaluation

In language education, testing, measurement, assessment and evaluation are interchangeably used terms. However, they do not refer to same things nor are they

synonymous. The basic difference may be made between the meaning of testing and assessment. While in testing, tests are used to measure language proficiency at one point of time, assessment is an ongoing process that takes into account the progress of language development (Brown, 2004). Evaluation, on the other hand, is a much wider term covering both testing and assessment procedures. Thus, it can be said that there are subsets of *evaluation* beginning with *test* in the core to *assessment* in the upper place (see Figure 2.1 for the relationship).

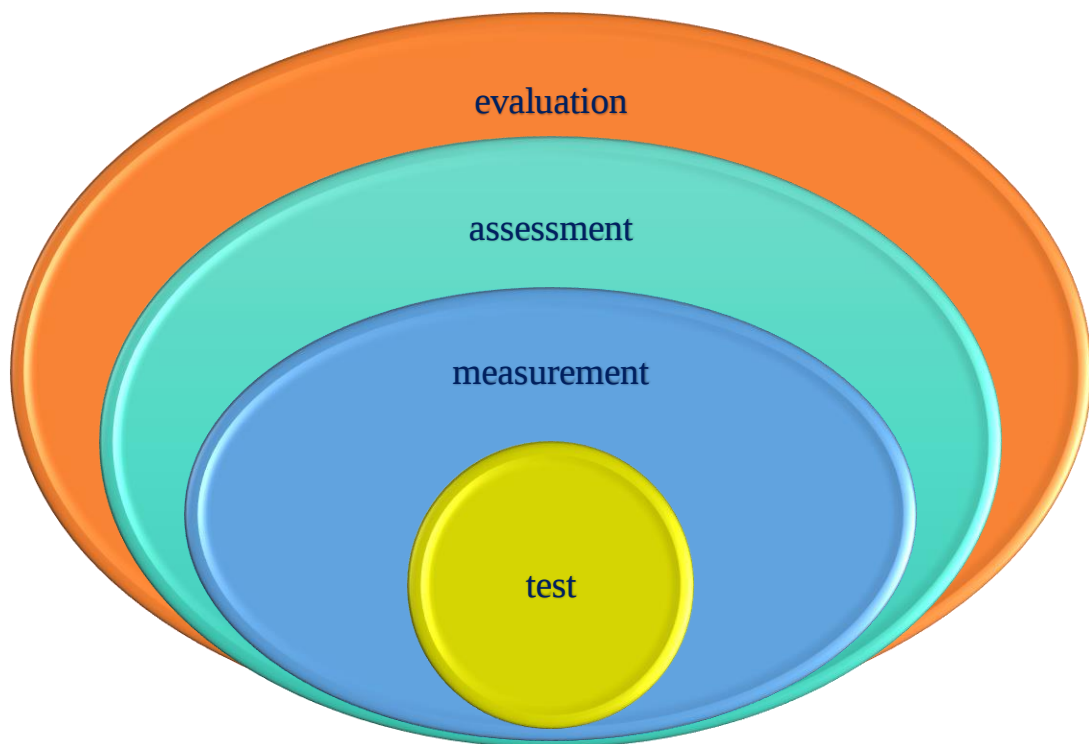


Figure 2.1. *The relationship among the basic terms*

2.2.1. Testing

The concept of testing can be best understood with the meanings of test and measurement. Besides, the purposes, types, principles and criteria of language testing may illuminate the meaning of testing more.

2.2.1.1. Definition of testing, test and measurement

Testing entails both uses of tests and measurement in order to find out to what extent language knowledge and skills are achieved by learners. In the literature, most researchers have described test, measurement and testing in a similar vein. For example, Coombe

(2018) has defined *testing* as ‘the process of seeing how we match up to a standard or standards’ (p. 41). On the other hand, *measurement* refers to ‘the practice of quantifying a physical feature or an attribute. [...] Assigning test scores to someone’s vocabulary knowledge or reading proficiency are both examples of measuring an attribute’ (Coombe, 2018, p. 29). Similarly, Bachman (2005) has defined *measurement* as ‘the process of quantifying the characteristics of an object of interest according to explicit rules and procedures’ (p.8). Hence, the concept of measurement is about quantity or numeric scales which classify knowledge into scores. As for the term of *test* itself, it means ‘a particular type of measurement that focuses on eliciting a specific sample of performance’ (Bachman, 2005, p.9). In other words, it refers to ‘a method of measuring a person’s ability; knowledge, or performance in a given domain’ (Brown, 2004, p. 3). Likewise, Coombe (2018) has indicated that *test* is ‘a set of tasks or activities intended to elicit samples of performance which can be marked or evaluated to provide feedback on a test taker’s ability or knowledge’ (p. 40). Thus, test is a method, way or type of measurement which is helpful to determine language knowledge and performance.

To sum up, it can be concluded that testing is a way of measuring the proficiency, ability or performance of language learners by using tests. In other words, tests are the tools to collect data about learners’ language proficiency, and measurement is the numerical display of how successful a language learner is. That is to say, this entire procedure that is carried out to measure language ability by using tests is called testing.

2.2.1.2. Purposes and types of tests and testing

In an educational context, language teachers or other stakeholders such as school principals or administrators make some decisions about language learners according to certain goals or objectives. In terms of testing language learners, the main purposes are placement, diagnosis, proficiency, achievement and aptitude (Brown, 2004; Heaton, 2011; Hughes, 2003). Accordingly, these purposes have led to certain types of tests to emerge.

Proficiency tests: These tests are ‘designed to measure people’s ability in a language regardless of any training’ (Hughes, 2003, p. 11). In the same vein, Heaton (2011) has stated that *proficiency tests* refers to ‘defining a student’s language proficiency with reference to a particular task which he or she will be required to perform; no related to any syllabus or teaching program; measuring simply students’ control of language for

future performance’ (pp. 172-173). Similarly, Brown (2004) has drawn attention to the characteristic of proficiency tests, and has argued that these types of tests are not based any syllabus or curriculum, but their purpose is testing overall language ability.

Achievement/attainment tests (Class progress and standardized tests): Contrary to proficiency tests, achievement tests are ‘directly related to language courses, their purpose to establish how successful individual students, groups of students or the courses themselves have been in achieving objectives’ (Hughes, 2003, p. 13). In addition, Heaton (2011) has noted that achievement tests are divided into: *Class progress* test which measures ‘the extent to which students have mastered the material taught in the class’, and *standardized test* that measures ‘achievement or attainment on a larger scale such as entrance exam’ (pp. 171-172). Brown (2004) has also defined *standardized tests* ‘are designed to apply to a broad band of competencies that are usually not exclusive to one particular curriculum’ (p. 67). Therefore, most of the time, achievement or attainment tests are interested in the product of learning; that is, summative side of language development (Brown, 2004).

Aptitude/prognostic tests: These tests are about the target language which has not been learned yet and are used to test future probable performance or ability in that language by the help of an artificial language (Heaton, 2011). Similarly, Brown (2004) has maintained that the main purpose of aptitude tests is to predict a learner’s success before learning a foreign or second language. He has explained that these tests are ‘designed to measure capacity or general ability to learn a foreign language and ultimate success in that undertaking’ (p. 43).

Placement tests: The main intention of these tests is to classify language learners in a teaching program according to their abilities or performances (Hughes, 2003). Likewise, Brown (2004) has illustrated that placement tests are used with the purposes of placing a language learner into the most appropriate language level of a school.

Diagnostic tests: These tests, as the name indicates, is ‘used to identify learners’ strengths and weaknesses, intended to ascertain what learning still needs to take place’ (Hughes, 2003, p. 15). Besides, Brown (2004) has pointed out that the aim of diagnostic tests is to analyze or identify certain characteristics or features of the target language that need to be concentrated on more. On the other hand, Heaton (2011) has argued that diagnosis just represents a purpose, and hence, these kinds of tests can be utilized as achievement or proficiency tests.

Other than these types of tests with different purposes, there are other categories with regard to language testing highlighted by some researchers. The classifications are based on the contrasts of two ideas or approaches: Formal vs informal; direct vs indirect; norm-referenced vs criterion-referenced, high-stakes vs low-stakes, discrete-point vs integrative, objective vs subjective testing, pen-and-paper vs computer testing.

Formal vs informal testing: While *formal testing* is measuring a language learner's ability, knowledge or performance within a course design as a sum of learning by using tests, *informal testing* is continuously evaluating a language learner's progress through implicit ways such as observation (Coombe, 2018). To exemplify this kind of testing, Spratt, Pulverness and Williams (2016) have noted some techniques. For instance, 'diagnostic, placement, progress, formative, continuous, portfolio, achievement, summative or proficiency' are examples of formal testing (pp. 104-105). In informal testing, there are testing or evaluating ways to be used both for teachers and learners such as observing, self or peer-assessment and feedback (Spratt, Pulverness & Williams, 2016).

Direct vs indirect testing: *Direct testing* is related to the correct use of language that is intended to be measured whereas *indirect testing* is interested in the underlying skills that are attempted to be measured (Hughes, 2003).

Norm-referenced vs criterion-referenced testing: In *norm-referenced testing*, each language learner is tested in relation to other language learners in the context and the acquired test score is interpreted related to percentile, rank, median, mean and the like (Brown, 2004). That is, norm-referenced testing 'relates one candidate's performance to that of other candidates' (Hughes, 2003, p. 20). On the other hand, *criterion-referenced testing* is used to grade a language learner's achievement in relation to specific course objectives (Brown, 2004). In other words, criterion-referenced testing 'is to classify people according to whether they are able to perform some task or set of tasks satisfactorily' (Hughes, 2003, p. 21).

Discrete-point vs integrative testing: While *discrete-point testing* means testing only one of language element at a time, *integrative testing* refers to testing more than one language element at a time, which requires test-takers to combine language components to fulfill the test (Hughes, 2003). For instance, testing only one grammatical structure item by item can be an example of discrete-point testing; cloze tests, however, requires

grammar, vocabulary and discourse knowledge at the same time, which is an example of integrative testing (Brown, 2004).

High-stakes vs low-stakes testing: *High-stakes* tests are generally external standardized tests which require testing experts to design and administer such tests due to their high impact and involving other stakeholders apart from learners such as educators or schools (Bachman & Palmer, 1996; Coombe, 2018). High-stakes exams are also called as *gatekeeping exam* because of its selective purposes (Coombe, 2018). Contrarily, *low-stakes testing* is based on low impact in that *classroom testing* is carried out and teachers test small group of learners (Bachman & Palmer, 1996). Especially in relation to low-stakes testing, another term ‘*teacher-made test*’ is used synonymously because this type of test ‘is designed by a teacher to measure a specific lesson’s objectives related to what is being taught in the classroom. This test can be written in more or less formal or informal ways, depending on the purpose of what the teacher is trying to achieve’ (Coombe, 2018, p. 40).

Objective vs subjective testing: Objectivity and subjectivity are only about the scoring of a test (Heaton, 2011; Hughes, 2003). *Objective testing* requires one correct answer for each question and is controlled by an answer key not including personal opinions and scored mechanically whereas *subjective testing* is much more about judgments of test-developers on test-takers’ test results and can be scored holistically (Heaton, 2011; Hughes, 2003). For instance, multiple-choice test and true/false items are examples of objective testing while oral production tasks and writing compositions are related to subjective scoring (Coombe, 2018).

Pen-and-paper vs computer testing: In *pen-and-paper tests*, test-takers respond to a paper by writing their answers, and it is mostly compared with either performance-based testing or computer-based testing (Coombe, 2018). However, computer-based testing is not a purpose, but it is rather a tool used to test language learners. In *computer-based testing*, test-takers connect to the Internet to be online or respond to questions in an offline program presented on computers (Coombe, 2018). Computer-based testing has also synonyms such as computer-assisted, computer-aided or computerized testing.

2.2.1.3. Principles or criteria of testing

Any formal language test requires some properties to be called as a good test. These kinds of properties are named as principles, qualities, criteria or key concepts of a good

test. That is, the qualities of good testing serve as test usefulness. Then, the principles or criteria of good testing are made up of reliability, validity, practicality, authenticity, washback (backwash/impact), interactiveness, and discrimination.

Test usefulness: Any test requires a balanced proportion of the qualities that make a test a good test such as reliability and validity (Bachman & Palmer, 1996). It also refers to the purposefulness, which is defined as any test needs to correspond to certain purposes of testing in mind (Rogier, 2014). For example, a teacher should consider what they teach at their lessons, and what kind of test needs to be administered to a certain group of students according to his/her testing goals. Also, such a developed test should contain the mentioned principles to be useful.

Reliability: *Reliability* is about the consistency and dependency of measurement or test scores (Bachman & Palmer, 1996; Brown, 2004; Coombe, 2018; Rogier, 2014). In other words, if any test is administered to the same group of language learners at different times, the scores the students get are expected to be consistent and similar (Bachman & Palmer, 1996; Brown, 2004; Heaton, 2011). Then, this test is called a reliable test. There are some forms of reliability for testing. For example, Brown (2004) has discussed student-related reliability, test reliability, test administration reliability, and rater reliability. *Student-related reliability* refers to students' physical (i.e., illness) or psychological (i.e., anxiety) problems that affect the reliability of test scores. *Rater reliability* is used against subjectivity and bias, and in order to prevent them, two raters can evaluate test scores and their comparison of test scores can be assessed whether the comparison is reliable between two raters' scoring. *Test administration reliability* is about testing conditions which can influence test-takers' performance and may yield different scores causing unreliable testing. *Test reliability* is about the nature of the test itself, i.e., if it is too long, if the allocated time is too short and if the test items are poorly written, then, the reliability of test scores are affected negatively, which decreases test reliability. In addition to Brown's (2004) discussion of rater-reliability, Coombe (2018) has mentioned inter-rater and intra-rater reliability. While *inter-rater reliability* is 'the level of agreement between two or more independent raters in their assessment of test takers' same performance' (Coombe, 2018, p. 25), *intra-rater reliability* is 'the extent to which a rater is internally consistent in using a rating scale by comparing two independent assessments of the same test performance by the same rater' (Coombe, 2018, p. 26). Apart from Brown's (2004) discussion, Heaton (2011) has argued other forms of reliability

directly related to test itself, namely, test reliability, such as test/re-test reliability, parallel forms and split-half method. *Test/re-test reliability* is administering the same test to the same group but at two different times; if the scores of two administrations are consistent, then, the test is reliable. The use of *parallel forms* is another reliability form which refers to the administration of two similar versions of tests to the same group, and if the scores got for two tests are similar, the test is said to be reliable. *Split-half method* is measuring test reliability through dividing a test into two parts and the scores got for each part are expected to be similar in order that the test becomes reliable. Heaton (2011) has also noted some factors that have an impact on the reliability of tests such as selected sample of material, administration, instructions, personal factors, and scoring, all of which may cause an increase or decrease in terms of reliability. Therefore, some researchers have suggested the ways to increase reliability of tests such as consisting of both easy and difficult items, being long enough to cover the content, giving not tricky questions and not trapping test-takers, clear and explicit instructions, writing ambiguous items, giving familiar formats that test-takers know, and providing a detailed scoring key (Hughes, 2003; Rogier, 2014).

Validity: *Validity* is the accuracy of measurement rather consistency. In other words, a test is valid if it measures accurately what it is intended to be measured (Brown, 2004; Coombe, 2018; Hughes, 2003; Rogier, 2014). Most researchers have explained more or less the same kinds of validity are required for a test to accurately measure what it is planned to be measured. For instance, Brown (2004) has underlined content-related evidence, criterion-related evidence including concurrent and predictive validity, construct-related evidence, consequential validity and face validity. Heaton (2011) has also exemplified face validity, construct validity and empirical validity. In the same vein, Hughes (2003) has noted content validity, criterion-related validity, construct validity, and face validity. *Content-related evidence (content validity)* refers to ‘the extent to which the items or tasks on a test constitute a representative sample of items or tasks of the knowledge or ability to be tested. In a classroom-teaching context, they are related to a syllabus or course’ (Coombe, 2018, p. 15). *Criterion-referenced evidence (criterion-related validity)* is about whether the determined criterion is reached or not. It has two types: *Concurrent validity* means ‘the correspondence between the scores of two different assessments measuring the same construct’ (Coombe, 2018, p. 15) whereas *predictive validity* is ‘the extent to which a test predicts performance in an external situation or future

performance’ (Coombe, 2018, p. 34). *Construct-related evidence (construct validity)* is about measuring not directly observable constructs whether it intends to be measured correctly such as proficiency and competence, and it also reflects a theory, model or hypothesis of language learning (Brown, 2004; Coombe, 2018). *Consequential validity* is about the influential consequences of tests on learners, teachers, society and the like in terms of interpretation of test scores (Brown, 2004; Coombe, 2018). *Empirical validity (statistical validity)* is the comparison of test results with another result of criterion such as an existing test, teachers’ ratings given later and subsequent performance of students (Heaton, 2011). *Face validity* is about the surface features of a test which make it look like a test to test-user as well as test-taker in term of whether it seems reliable, valid, fair, acceptable and free of ambiguities (Brown, 2004; Coombe, 2018; Heaton, 2011). Besides, certain factors can influence the validity of tests such as unclear instructions, asking unknown parts of language which are not studied in the course, poorly written items, insufficient coverage of content and complexity of test (Rogier, 2014). Thus, some researchers have recommended how to increase validity such as constructing test specifications as a table, including representative samples of the content, using direct testing, and scoring in line with the measured construct (Hughes, 2003; Rogier, 2014).

Practicality: *Practicality* refers to the administrative details or available and required resources of a test such as time, cost, materials and scoring (Bachman & Palmer, 1996; Brown, 2004). Furthermore, Rogier (2014) has pointed out that test practicality includes ‘the cost of test development and maintenance, time needed to administer and mark the test, ease of marking, availability of suitably trained markers, and administration logistics’ (p. 5). Likewise, Coombe (2018) has argued that practicality concerns such issues as time constraints, space, administrative staff, resources used, preferences and policies. All of the discussed points are about feasible and applicable side of testing.

Authenticity: *Authenticity* is about to what extent a test represents real-life tasks, contexts and situations in terms of language use (Coombe, 2018; Rogier, 2014). Thus, the characteristics of testing items or tasks need to reflect appropriate and relevant real-life uses of the target language (Brown, 2004). If it did not reflect, then, it would be useless and not meaningful to test-takers.

Washback (Backwash/Impact): As the name reveals, *washback* is about the relationship between testing and teaching, and the effect of testing practice on teaching and learning of a language (Brown, 2004). It may be either positive, beneficial, motivating

or negative, adverse, deteriorating, harmful. For example, *positive impact* occurs when test-taker is familiar with the given test task type and topic, and easily perform it whereas *negative impact* happens when test-takers get adverse feedback and do not receive positive responses about their test results (Bachman & Palmer, 1996; Coombe, 2018). Moreover, the impact of testing is not only on teaching but also on test-takers, teachers, schools, curriculum, educational systems, parents, society and the like either at micro or macro level, all of which are more or less affected by test results positively or negatively (Bachman & Palmer, 1996). Specifically, external, high-stakes and standardized-tests have been found to affect test-takers negatively when the expectations are very high and test-takers' only intention to pass the test rather than learning to use the target language communicatively (Brown, 2004; Coombe, 2018). Besides, there are some ways to sustain beneficial washback. For instance, encouraging a certain language component to improve by testing it, making test specifications clear and explicit to test-takers, using direct testing with authentic materials, basing achievement tests on the course objectives rather than the coursebook, ensuring the test comprehensibility and familiarity, providing assistance, providing constructive feedback on the results and involving test-takers into scoring process (Hughes, 2003; Rogier, 2014). To sum, washback shows that a test can influence 'what teachers teach, how teachers teach, what learners learn, how learners learn, the rate and sequence of teaching, the rate and sequence of learning, attitudes to the content, method, etc. of teaching and learning' in a classroom setting (Fulcher, 2010, p. 277).

Interactiveness: *Interactiveness* is one of the qualities of a good test which refers to 'the extent and type of involvement of the test taker's individual characteristics in accomplishing a test task' (Bachman & Palmer, 1996, p. 25). That is, individual characteristics of test-takers influence their interaction with the test task. For instance, topical knowledge, affective domain, strategic competence and metacognitive strategies of a test-taker have an impact on his/her performance in terms of whether interacting with the test appropriately (Bachman & Palmer, 1996).

Discrimination: *Discrimination* refers to the capacity of a test to what extent it discriminates among different test-takers' abilities (Heaton, 2011). Specifically, it is about differentiating test-takers' ability according to handling easy and difficult test items, which is useful in terms of discriminating successful students from low-ability students (Coombe, 2018).

2.2.2. Assessment

The concept of assessment is a broader term compared to the concept of testing. Its meaning can be best understood through its definitions and types used in language testing and assessment.

2.2.2.1. *Definition of assessment*

Assessment can be seen as a different concept than testing because it refers to a whole process of gathering data about language progress by using different kinds of techniques and methods in order to find out what is learned, what needs to be fulfilled, and how to improve learning and teaching effectiveness. Clapham (2000) has mentioned that assessment has two underlying meanings: A general term for all methods of testing and assessment, and a specific term used to differentiate alternative assessment from testing. Moreover, Clapham (2000) has illustrated that some researchers use the term of testing to refer to the development and administration of tests whereas others indicate the term of assessment to mean using more informal methods rather than tests such as school projects, interviews, portfolios and demonstrations. Similarly, Bachman (2005) has exemplified the methods for assessment such as elicitation tasks, observation, essays, portfolios, and even standardized tests and multiple choice tests, all of which can be utilized to draw information about a language learner's ability or progress, and the results can be reported both quantitatively and qualitatively. Furthermore, Coombe (2018) has underlined the issue of systematicity in the definition of assessment, and he has described it as 'the systematic process of evaluating and measuring collected data and information on students' language knowledge, understanding, and ability in order to improve their language learning and development' (p. 10). Lastly, Spratt, Pulverness and Williams (2016) have drawn attention to the concept of judgment while discussing the definition of assessment, and they have stated that assessment means 'judging learners' performance by collecting information about it [...] for different reasons and using different methods' (p. 104).

In conclusion, assessment is a cover term both for testing procedures such as designing and administering, and for techniques, methods or tools such as tests, oral tasks, portfolios and projects used to gather information about a language learner's ability, knowledge and performance regarding the use of the target language.

2.2.2.2. *Types of assessment*

Language assessment practices have been shaped in line with teaching methodologies and perspectives over the years. There has been a shift from more traditional educational practices to more communicative and localized language teaching applications, and similarly, in language assessment. For instance, Campbell and Duncan (2007) have argued that there has been emphasis on communicative language teaching, learner-centered instruction, content and task-based instruction, and learner autonomy in terms of language assessment from 1970s to 1990s. Afterwards, they have discussed that standards as proficiency guidelines for assessment, alternative assessment and dynamic assessment have been more influential in language testing and assessment from 1990s to date. In other words, within the years, there have been changes from *traditional, externally set, large-scale, standardized, formal exams (testing perspective) to teacher-based assessment, alternative assessment, formative assessment, assessment for learning, (assessment perspective)*’ in foreign language assessment. The assessment perspective is constructed through the view of ‘teacher-mediated, context-based, classroom-embedded assessment practice’ (Davison & Leung, 2009, p. 395). However, although there are a number of assessment types and techniques, most teachers or schools can decide and implement their language teaching as well as assessment practices according to either their content or needs of students or nationally accepted goals of language education.

Formal vs informal assessment: While *formal assessment* is carried out through formal procedures such as exams and tests in a systematic and planned way, *informal assessment* is conducted through classroom tasks such as observation and comments based on judgments spontaneously (Brown, 2004; Coombe, 2018; McKay, 2008). The purpose of formal assessment is to evaluate students’ achievement whereas informal assessment is integrated into classroom tasks to derive language performance without scoring or grading (Brown, 2004). The common types of formal assessment are written tests, standard-based performance test, diagnostic tests, placement tests, proficiency tests, and so on (Coombe, 2018; Spratt, Pulverness & Williams, 2016). As for informal assessment, incidental responses, comments, unplanned feedback, classwork, observation, self and peer-assessment are typical examples (Brown, 2004; Spratt, Pulverness & Williams, 2016). It should be noted that ‘all tests are formal assessments, but not all formal assessment is testing’ (Brown, 2004, pp. 5-6).

Summative vs formative assessment: The comparison between summative and formative assessment has been widely accepted as the main distinction in language assessment practices. Most of assessment practices or procedures are somehow part of either summative or formative assessment. While *summative assessment* refers to the evaluation of the sum of language learning at the end of the term, course or unit, *formative assessment* means evaluating the progress of language learning throughout the term, course or program more than one time (Brown, 2004; Coombe, 2018; McKay, 2008). So, it can be inferred that the *product* of language learning is important in summative assessment whereas the *process* of learning is crucial in formative assessment. That is, the aim of summative assessment is to find out how much students learn the target language at the end of the term; on the other hand, the purpose of formative assessment is to evaluate learning process so as to give feedback, to improve learning, and to make students realize their strengths and weaknesses (Coombe, 2018; Popham, 2009). The common techniques for summative assessment are course tests, final projects, final exams and proficiency exams (Brown, 2004; Coombe, 2018). Formative assessment procedures are made up of ongoing assessment procedures such as feedback, observation, portfolio and students' progress folders (Coombe, 2018). Then, in a sense, summative assessment contains more formal, and formative assessment comprises more informal practices.

Assessment of learning (AoL) vs assessment for learning (AfL): Assessment of learning (AoL) is actually the other name of summative assessment whereas assessment for learning (AfL) is synonymous with formative assessment. *Assessment of learning (AoL)* refers to 'a measurement or set of strategies that provides evidence of whether the learner has successfully achieved the learning objective of a particular program [...] where the information generated is not then used formatively to help learning' (Coombe, 2018, p. 11). On the other hand, *assessment for learning (AfL)* is the evaluation of language learning process to give feedback and improve learning and performance (Coombe, 2018).

Learning-oriented (language) assessment (Lo(L)A): *Learning-oriented assessment* seeks to find out how much language learners advance in the language, and if they have weaker points, language teachers can cover them by helping them to develop more and meeting their needs in the language (Carless, 2015; Hamp-Lyons, 2017). This type of assessment can be either summative or formative as long as it provides the development of learning (Carless, 2007; Green, 2016). Thus, the main assumption behind

learning-oriented assessment is to enhance language learning. Learning-oriented language assessment can be achieved through learners themselves as self or peer assessment, teachers as doing formative assessment, and external staff or teachers doing summative assessment (Green, 2016). In addition, learning-oriented assessment tasks should include the characteristics of learning activities, student involvement or engagement, and promoting learning with feedback (Carless, 2007). Therefore, as the name suggests, learning is the central assessed construct in order to foster language learning itself.

Classroom assessment: *Classroom assessment* is executed by teachers so as to make inferences about students' language progress, and teachers can utilize both formal and informal procedures (Popham, 2009). It is one of the formative assessments and assessment for learning ways because the main aim is to evaluate language learning process. So, classroom assessment provides teachers with the feedback about the effectiveness of their teaching, and it helps students to make more progress in language learning (Coombe, 2018). In contrast to external assessment which is carried out by the assessors outside the class, classroom assessment is both designed, monitored and performed by teachers (McKay, 2008).

Accountability assessment: Being one of the formative assessment types, *accountability assessment* refers to 'measurement devices, almost always standardized, used by governmental entities such as states, provinces, or school districts to ascertain the effectiveness of educational endeavors' (Popham, 2009, p. 6). They are used for the purposes of accountability; that is, to what extent whether teachers, students, program, institution or system are successful in providing academic achievement, and how much they are able to inform others about the performance as well as development of language learning.

Teacher-based assessment (TBA): Teacher-based assessment are used interchangeably with different terms such as classroom assessment, school-based assessment, formative assessment, and assessment for learning (Davison & Leung, 2009). It refers to teacher-made and teacher-conducted assessment, and thus, non-standardized assessment but classroom-based assessment (Y. Xu, 2016). While planning such assessment, teachers need to consider the answers to the questions of who, what, where, why, and how as in the stages of developing and administering language tests in the classroom (Davison & Leung, 2009).

Continuous assessment: Being one of the formative classroom assessments, *continuous assessment* refers to evaluating students' language progress and controlling periodically whether learning objectives are achieved (Coombe, 2018). In order to perform continuous assessment, language teachers can make use of in-class tests, portfolios, presentations, quizzes, assignments and projects (Coombe, 2018).

Traditional assessment vs alternative assessment: While *traditional assessment* is about one-shot, standardized, summative, product-oriented, and non-interactive performance, *alternative assessment* refers to continuous, process-oriented, formative, open-ended, free-response, and interactive performance (Brown, 2004). Therefore, traditional assessment is interested in the sum of learning through end-of-term exams whereas alternative assessment is concerned about the development of learning during the process. Generally, traditional assessment utilizes pen-and-paper tests, short-answer tests and multiple-choice tests for the purposes of recalling and recognition. However, alternative assessment employ projects, essays, journals, interviews, conferences, portfolios, observations and the like in order to track students' language development and to make them to use or produce language in different ways (Brown, 2004; Coombe, 2018; Uyaniker, 2017). There are several types of alternative assessments reported in the literature, but the popular ones are performance-based assessment, portfolio assessment, self-assessment, and peer-assessment.

- **Performance-based assessment:** The fundamental aim of *performance-based assessment* is to assess the performance of language learners in the target language. Therefore, productive and observable skills are assessed through authentic and real-life tasks (Brown, 2004). The common techniques used for performance-based assessment are projects, demonstrations, presentations, and role-play (Coombe, 2018).

- **Portfolio assessment:** *Portfolio* refers to 'a purposeful collection of students' works that demonstrates their efforts, progress and achievements in given areas' (Brown, 2004, p. 256). They can be collected via folders and digital platforms to document students' progress in the language (Coombe, 2018). The popular contents of portfolios consist of essays, compositions, drafts, reports, projects and recordings (Brown, 2004; Vogt & Froehlich, 2018a).

- **Self-assessment:** *Self-assessment* means 'the ability of language learners to assess their own performance to identify their strengths and weaknesses in the learning process' (Coombe, 2018, p. 37). Language learners evaluate their development on their

own by using for example a kind of checklist. Therefore, self-assessment is autonomous, and helpful to increase intrinsic motivation (Brown, 2004; Coombe, 2018).

- **Peer-assessment:** *Peer-assessment* is conducted by other students as peers through assessing each other's works by means of a criterion either provided by teacher or determined by learners (Coombe, 2018). It is based on the idea of cooperative learning because pairs or group of learners assess each other's work while making one another think critically about their language progress (Brown, 2004; Coombe, 2018).

Dynamic assessment (DA): One of the recent assessment approaches is dynamic assessment, which has been dominantly discussed by Lantolf and Poehner (2011) within sociocultural theory of Zone of Proximal Development (ZPD) by Vygotsky. It is based on the view that assessment and teaching are inseparable, and cognitive abilities can be mediated and developed through social interactions (Fulcher, 2010). It is not a specific assessment method or technique, rather it is a conceptualization to understand learner abilities (Lantolf & Poehner, 2011). *Dynamic assessment* is used to evaluate what learners can do on their own (about their potential language learning) and what they can do when supported and mediated by their teachers or peers (Uyaniker, 2017). That is, the difference or interaction between the independent performance of language learners and guided performance are assessed to improve language learning.

2.2.3. Evaluation

Evaluation can be seen as the last step but the covering term for assessment findings and test results. It is about scoring, grading and marking the results and accordingly, giving feedback to learners as well as to teachers to improve the development of language learners and the effectiveness of teaching. Bachman (2005) has argued that evaluation 'involves making value judgments, and decisions, can best be understood as one possible use of assessment. For example, use information from assessment to make decisions about selection and placement and to assign grades or marks' (p. 9). Similarly, Coombe (2018) has illustrated that evaluation refers to 'the process of utilizing assessment results to gauge and support learning, teaching and program development' (p. 20). Therefore, evaluation can be defined as the process of collecting data and interpreting results systematically in order to make decisions purposefully.

With regard to language testing, evaluation can be made through either grading as numerical or letter grading or through other alternatives. For example, Heaton (2011) has

reported that while interpreting test scores, some numerical values such as frequency distribution of the sample, mode, median, range, item difficulty of the test and the like are calculated to help making decisions about students' success and failures. Furthermore, Brown (2004) has indicated that distribution of scores are graded as numerically to show whether students are good or bad. For instance, a student takes 90 points from a grammar test, and if the highest test score is 100 points, then, this student can be evaluated as being successful. In addition to numerical grading, teachers can give letters to the distribution of scores; i.e., A = excellent, B = good, C = adequate, and so on. Apart from such grading techniques for student evaluation, there are other ways. For instance, self-assessment comments, narrative evaluations, checklist evaluations and conferencing with students about their progress (Brown, 2004).

However, evaluation is not limited to only classroom or student decisions but also the programs, syllabi or administrators can be evaluated to make decisions about their effectiveness. For instance, Bachman and Palmer (1996) have discussed that evaluation requires the decision-making in relation to test-taker, teacher or administrator, and programs. While test-takers are tested with the purposes of placement, diagnosis or progress according to language development, teachers or administrators are tested about the effectiveness, retention or promotion. Pertaining to programs, they are tested whether they are good, need modification or elimination. In the same vein, Bachman (2005) has argued that there are two types of decisions made with the aim of evaluating: Decisions about individuals such as selection, placement and diagnosis, and decisions about programs such as modifying, eliminating or continuing the existing program. In this sense, Coombe (2018) has defined a specific term *program evaluation* as 'an assessment of an instructional program, such as a language school or a degree-granting unit within an academic institution' (p. 34).

To conclude, evaluation is an umbrella term dealing with making purposeful decisions through scoring, grading, interpreting and feedback in terms of learners, teachers, syllabi and programs.

2.2.4. Other associated terms and concepts

Recognition vs production: *Recognition* is about receptive side of testing in which test-takers notice the correct forms or words in a language test (Heaton, 2011). For example, in a multiple-choice test, a test-taker chooses the correct answer by looking at

the given choices according to the question or statement. On the other, *production* is about language use and construction by writing or speaking to show test-takers' ability (Heaton, 2011). For instance, completing a dialog in a test requires test-takers to produce some forms of language.

Feedback: *Feedback* is about the evaluation and critique of the test result or task performance of a language learner given either by his/her teacher or peers in the class (Coombe, 2018). If the feedback is positive and constructive, then, language learners will be motivated to learn more and make them to see their strengths and weaknesses. In addition, feedback can be written, spoken or just a symbol to motivate learners more (Vogt & Froehlich, 2018b).

Marking (grading/scoring): *Marking* refers to 'the process of assigning numerical or otherwise systematic values to the performance of a candidate in an assessment' (Coombe, 2018, p. 28). For example, it can be conducted by comparing test-takers' results with the answer key whether they are matched.

Rater. *Rater* is an individual evaluator whose task is to assess test-takers' spoken or written performance related to their language use (Coombe, 2018).

Rubric (Rating scale): *Rubric* is a type of scale or grid which has numerical values to evaluate test-takers' spoken or written language performance in a test (Coombe, 2018). There are two forms of rating scales used to assess language performance: *Holistic or impressionistic marking* is evaluating language performance through the opinions of the rater without definite descriptors whereas *analytic scale* utilizes certain descriptors such as topic, coherence, cohesion and language use that each of them has definite numerical values to calculate the performance of tests-takers (Coombe, 2018).

Benchmark: *Benchmark* is a kind of reference with which the measured language element is compared and contrasted. In other words, it is 'a standard or a point of reference against which learning is measured or compared. It is the reference point that you measure mastery, achievement or progress against' (Coombe, 2018, p. 12).

Bias: *Bias* is a kind of opinion or judgment about the language test content which requires other types of knowledge such as culture to comprehend it and if such knowledge did exist, the test would be called as a biased test. That is, a test can be biased 'when the test-taker has to have some assumed cultural knowledge outside the test to understand the question or the text' (Coombe, 2018, p. 12).

Ethics: This concept refers to ‘a system of established and acknowledged principles that guide the construction, development, administration, management and evaluation and use of language assessments in any context’ (Coombe, 2018, p. 19). The most popular one is *ILTA Code of Ethics* which has nine principles about ethics in language testing. For example, language testers should respect for humanity and dignity or they should not allow misuse of knowledge.

Fairness: *Fairness* is about both validity and reliability. It refers to ‘validity during test construction, consistency of implementation during test administration, and consequences for use of test scores for individuals or groups, examiners’ judgments or evaluations, and society at large’ (Coombe, 2018, p. 21).

Transparency: *Transparency* means being clear and explicit about test uses and measures. That is, it refers to ‘making information about the assessment process and criteria accessible and available to users, including what the measures are. For example, the use of rubrics is a way of providing transparency in assessment’ (Coombe, 2018, p. 42). In short, it is the indicator of the availability of test information to test-takers by making them to understand goals of lessons, the ways of scoring and grading, and the content of what is tested (Rogier, 2014).

Needs analysis: *Needs analysis* refers to the collection of data about the needs of language learners in order to plan what to teach as well as what to test (Coombe, 2018). Such analysis yields better educational opportunities for students to improve their target language knowledge and performance because lessons, course, syllabus or curriculum are constructed in terms of responding to their needs.

Stakeholder: In education, those whose needs, interests and opinions are taken into consideration while developing test, syllabus, course or curriculum are the stakeholders because those individuals are affected by and also affect the decisions about the design as well as the consequences (Coombe, 2018; McKay, 2008). There are several stakeholders that should be considered in testing; i.e., teachers, students, parents, administrators, community, faculty and staff, raters, schools and authorities.

Test anxiety: *Test anxiety* refers to psychological factors that affect a test-taker’s ability or performance in language testing negatively. For example, stress, worry, fear of failure and nervousness may cause a decrease in a test-taker’s performance (Coombe, 2018).

Test battery: *Test battery* is a group or set of tests for specific purposes to measure specific language components (Coombe, 2018). For example, testing reading, writing, listening and speaking generate a basis to measure language proficiency.

Target language use domain (TLU): This concept refers to the purposeful use of target language in communicative tasks, which has a role in conceptualization because it shows test-takers to use language except from the test itself. That is, *TLU* means ‘goal-oriented contexts of language use in which language users perform communicative tasks [...] the context of actual language use in which a learner will be using the target language outside the test itself’ (Coombe, 2018, p. 39).

CEFR: *Common European Frame of Reference* is a framework which describes language proficiency according to certain criteria in order to represent the level of language learners in that section. It has ‘can do’ statements which indicate what language learners are able to do in a specific language level; i.e., A1 is basic while C2 is proficient user of language (Coombe, 2018). It can be utilized with different purposes such as achievement and placement while providing a standard comparison across European countries.

2.2.5. Foreign language testing, assessment and evaluation

Testing, assessment and evaluation are important aspects in foreign language education, especially at the classroom level, because it is useful to improve and monitor the progress of language learners while they are learning a new language which is spoken or not spoken outside the class. In this sense, there is a need to differentiate between *second language assessment* and *foreign language assessment*. While the progress in acquiring a second language can be more fruitful because the target language is used outside the class in everyday communication and assessment can be used more with the aim of improving the language abilities and skills, the development in learning a foreign language is too difficult since the target language is not spoken outside the class and sometimes the class is the only context where they can use the target language except from using it on the Internet or another context. Therefore, foreign language assessment is necessary to monitor the progress, feed their language performance, and to more motivate language learners to learn the target language.

In general, language assessment is used to collect information about language competence and performance of language learners; so, it is helpful to draw information,

and accordingly, make inferences about the progress of language learners (Djoub, 2017). In terms of testing perspective, Heaton (2011) has stated that a language test presents how successful language learners in making meaning; that is, using the target language purposefully. Besides, language assessment can be useful in terms of its functions as being summative (end-of-term evaluation), formative (evaluating strengths and weaknesses during the learning process), and integrative (making teachers plan their further teaching practices).

In fact, in language assessment, language proficiency is tested, assessed and evaluated. In the simplest terms, *language proficiency* refers to the possession of adequate language knowledge such as grammar, vocabulary and pronunciation, and the use of language skills such as listening, speaking, reading and writing effectively (Bachman & Palmer, 1996). In addition, the construct of language ability requires not just language knowledge but also topical knowledge, strategic competence and personal features to be able to use language efficiently (Hidri, 2018). Thus, observation and performance are important points to be considered while assessing language ability; i.e., teachers assess competence whereas they observe performance (Brown, 2004).

Language teachers are the central figures in order to assess language proficiency of learners because they spend most of their time on teaching the target language, and they usually take care of their students' learning on their own in the classroom. Also, language teachers often make use of classroom tests and assessments to monitor the learning process, and generally, most teachers prepare their own language testing and assessing tools according to their course objectives as well as students' characteristics and needs. This type of testing or assessment is called *teacher-based assessment*, which reflects localized testing and assessment procedures rather than standardized-testing, and mostly used with the aim of improving learning (Davison & Leung, 2006; Y. Xu, 2016). Then, teacher-based classroom assessment is very important to enhance language proficiency of learners. However, such assessment may be affected by such factors as beliefs of teachers, local needs, nature of setting, and goals of learning (Munoz, Palacio & Escobar, 2012). Apart from these factors, the knowledge and skills of preparing and administering language testing and assessment as well as evaluating language learners' proficiency are the other factors that might an impact on teachers' assessment practices. Therefore, language teachers need to have the necessary knowledge how to perform effective assessments since effective or sound assessment requires to be 'purposive and targets

clear and relevant objectives which can contribute to both evaluating and developing learners' language ability' (Djoub, 2017, p. 10). This kind of knowledge includes knowing how to design, administer and evaluate assessment according to language areas and skills.

Regarding the step of designing testing tools and assessments, first of all, they need to decide what to assess, when to assess, how to assess, and how to use the feedback to improve teaching and support learning. After that, they need to consider how much their means of assessment is reliable (i.e., grading or scoring), valid (i.e., aims and objectives) and practical (i.e., designing and administering) in their contexts. In this issue, Fulcher (2010) has focused on the criteria of selecting and using language tests; these are 'test purpose, test taker characteristics, domain of interest, constructs of interest, reliability, validity, parallel or equated forms, test administration and practicality, and impact' (pp. 292-294).

As for the stages of developing tests and assessments, some researchers have suggested certain ways to follow. For instance, Bachman and Palmer (1996) have illustrated that language test development has three stages as *designing*, *operating* and *administering* tests. In the stage of designing, it is important to describe the features of language tasks and purposes. In the stage of operationalization, task specifications are developed, language tasks that are intended to be measured are written, the instructions to the questions and tasks as well as scoring criteria are arranged. In the last stage, administration, the constructed language test is applied, the performance results are collected and evaluated. Similarly, Hughes (2003) has indicated that test development phases consist of 'stating the problem, constructing specifications (content, timing, scoring, etc.), writing and modifying items, test trial, calibration of scales, validation, writing handbooks for test users, and training staff' (p. 58). Likewise, Brown (2004) has demonstrated that while designing language tests, there are five steps to follow: Determining testing purpose, setting objectives, writing test specifications, selecting and organizing test tasks and items, scoring, grading and feedback. In brief, the researchers have pointed out a language test should be developed in line with certain criteria. Among the steps of designing language tests, writing *test specifications*, also named as the *blueprint*, is acknowledged to be the essential part because it increases the probability of making a more reliable and valid test as well as showing the content of testing to both students and teachers (Rogier, 2014). While general test specifications contain the

content, structure, timing, techniques, performance levels, scoring ways, reporting type, and number of tasks, an exam blueprint for each language skill that illustrates their weights in a language test (i.e., in percentages) by corresponding to the learning outcomes and objectives of the course (Rogier, 2014). Thus, the entire procedure about designing language tests require certain principles to be followed in order to construct reliable, valid, practical and appropriate tests.

Other than the stages of language test development, another topic that should be taken into account is common test types and techniques used in the classroom so that language teachers become familiar with and practice such methods in their classes. Language teachers make use of tests or assessments for the purposes of achievement, diagnostic or placement most of the time (Tuzcu-Eken, 2016). Besides, there are common techniques that can be employed in language testing such as multiple-choice items, yes/no questions, true/false statements, short answer questions, matching and gap filling tasks (Hughes, 2003). In addition to language testing, there are alternative ways to assess language learners such as performance-based assessment, portfolios, journals, conferences, interviews, observations, and self/peer assessment (Brown, 2004), all of which can be used by language learners to monitor language progress. However, there are other ways of testing and assessment for each language skill (listening, writing, speaking, reading) and area (grammar, vocabulary) since language components require different knowledge domains in correspondence to the level and age of language learners.

2.2.5.1. Assessing grammar

Grammar is one of the important language areas that is assessed separately on their own or integrated with other knowledge and skills. In assessing grammar, the purpose is to assess *grammatical knowledge (form and meaning)* as well as *pragmatical knowledge (use)* (Purpura, 2005). Hence, there is a grammatical ability to be assessed by language teachers which requires knowledge of language (form, meaning, use), strategic competence and metalinguistic knowledge (Purpura, 2005). That is, while testing grammatical ability of language learners, ‘phonological, graphological, lexical, morphosyntactic, cohesive, information management, interactional form and meaning’ are assessed (Purpura, 2005, p. 62). In short, grammatical ability refers to ‘the capacity to realize grammatical knowledge accurately and meaningfully in test-taking or other language-use contexts’ (Purpura, 2005, p. 89). Therefore, while designing grammar tests,

all of the elements of grammatical ability should be taken into consideration, and also, the purpose, objectives, target language use domain, tests tasks, timing, instruction, setting, structure, expected outcomes and scoring should be determined and organized. In terms of grammar test tasks, there are two types: objectively scored (i.e., multiple choice) and subjectively evaluated (i.e., essay) tasks. However, most common grammar test tasks are multiple-choice items, matching items, discrimination items, noticing, gap filling, short-answer, error correction/recognition, broken sentence items, combining or adding items, transformation exercises, changing of words, dialog completion, discourse completion task, extended production task, info-gap task, story-telling and reporting, role-play, and simulation task, all of which can be utilized to assess grammatical knowledge and ability (Heaton, 2011; Purpura, 2005; Spratt, Pulverness & Williams, 2016).

2.2.5.2. Assessing vocabulary

Vocabulary is words of a language and units of meaning that construct phrases, sentences, paragraphs and texts. Thus, assessing vocabulary means to assess word meanings and forms, sentence patterns and the like (Read, 2000). From this point of view, vocabulary ability consists of three elements: ‘*Context* of vocabulary use (used in sentence or utterance), vocabulary *knowledge* and fundamental processes (size, word characteristics, lexicon organization) and *metacognitive strategies* for vocabulary use (paraphrase, switch, superordinate terms, appeal to authority)’ (Read, 2000, p. 29). In the same vein, Heaton (2011) has indicated that there are two types of vocabulary: *Active* and *passive*, which are assessed in terms of word formation, word meaning and collocation. Thus, all of these elements should be considered while designing vocabulary tests. In the stage of developing vocabulary items, in addition to the decision of which elements of vocabulary is assessed, the purpose, the type – receptive (recognition and recall) or productive (comprehension and use), test input, presentation of words and characteristics of expected response should be taken into account. Specifically, it is important to consider which words are tested and how they are assessed (discrete or embedded, selective or comprehensive, context-dependent vs context-independent) (Read, 2000). As for the most common vocabulary test items, there are matching, completion, sentence writing, generic test items, word-associates, labeling, categorizing, word-building, word maps, odd one out, synonym/antonym, lexical sets, definitions, fill-in-the-blanks, and multiple-choice items (Heaton, 2011; Read, 2000; Spratt, Pulverness & Williams, 2016).

2.2.5.3. Assessing reading

Reading is about ‘perceiving the written form of language visually or kinesthetically’ (Alderson, 2001, p. 13). It is a *receptive skill* that cannot be observed, only reading comprehension performance can be assessed (Brown, 2004). Though it is a receptive skill, readers engage actively in two processes, *top-down* (background knowledge is used to predict meaning) and *bottom-up* (visual input as letters, words, sentences are used to check prediction), to comprehend the text (Csépes & Fekete, 2018a). Therefore, perceptive, selective, interactive, extensive reading tasks can be used while assessing reading comprehension (Brown, 2004). On the contrary, Csépes and Fekete (2018a) have argued that both receptive and productive tasks can be utilized in assessing reading comprehension. Receptive tasks can be exemplified as multiple-choice, true/false, matching or sequencing, and productive tasks are short-answer questions, gap-filling, information-transfer and summary writing. Hence, in assessing reading, actually ‘the ability to understand the gist of a text and to extract key information on specific points in the text’ is assessed (Heaton, 2011, p. 8). While designing reading tasks, the purpose, level, construct, time, target language situation, the layout, instructions, test length, text complexity/difficulty, test type, task type, number of items, test methods and criteria of scoring need to be arranged (Alderson, 2001). The commonly used techniques to assess reading comprehension are cloze test, gap-filling, multiple-choice, matching, ordering, short answer, free-recall, summary, gapped summary, information transfer, yes/no questions, open comprehension questions, choosing titles, skimming and scanning (Alderson, 2001; Heaton, 2011; Spratt, Pulverness & Williams, 2016).

2.2.5.4. Assessing listening

Listening is ‘a complex process in which the listener takes the incoming data, an acoustic signal, and interprets it based on a variety of linguistic and non-linguistic knowledge’ (Buck, 2002, p. 247). It is a *receptive skill* and performance in listening comprehension is assessed through intensive, responsive, selective, extensive and authentic listening tasks (Brown, 2004; Green, 2018). Accordingly, there are three approaches to listening test tasks: *Discrete point* (i.e., phonemic discrimination), *integrative* (i.e., cloze, gap filling), and *communicative* (i.e., authentic) (Buck, 2002). Thus, language teachers consider such points in designing their listening comprehension tasks because both text and task characteristics make it difficult or easy to comprehend

listening. The recordings used in assessing listening need to reflect spontaneous spoken language rather than reading written material in order to make students grasp the meaning from the speech while listening (Green, 2018). Language teachers also need to pay attention to the physical setting of testing, time, instructions, scoring, input, output, and task interactiveness while creating listening tasks to assess learners' ability to comprehend auditory signals (Buck, 2002). Common techniques or tasks used to assess listening comprehension involve comprehension questions, short-answer questions, multiple-choice questions, true/false questions, inference questions, gap-filling, tick the word/sentence heard, choosing the best answer/picture/etc. according to the description heard, dictation, public announcements, discussion, lectures, conversations, dialogs, drama, show, song, interview, commentary, and news (Buck, 2002; Green, 2018; Heaton, 2011; Spratt, Pulverness & Williams, 2016).

2.2.5.5. Assessing speaking

Speaking is seen as 'the linguistic descriptions of spoken language, interaction and social and situation-based activity' (Luoma, 2009, p. 9). It is a *productive skill* and language use of learners can be directly observed as their oral proficiency via imitative, intensive, responsive, interactive, extensive speaking/oral tasks (Brown, 2004). While assessing speaking ability, the use of language functions in real-life situations are evaluated. In other words, language teachers need to concentrate on 'the ability to use the language in lifelike situations, in which they have to perform a variety of language functions' while assessing oral production (Csépes & Fekete, 2018b). Moreover, not just speaking task but also other factors such as learners' characteristics, interlocutors' approach, topical knowledge and the number of participants assessed at one time can affect the performance of language learners in their oral production. In the classroom setting, there are three ways to assess learners' speaking: *Individual mode* in which teacher makes the assessment; *paired mode* in which two learners interact; *group mode* in which more than two learners interact (Csépes & Fekete, 2018b). While designing oral tasks, there are some points to be considered such as the context, quality of instruction, test mode, nature of task, live or tape-based task, task difficulty, quality of visual prompts or pictures, amount of participant contribution, task specifications and scoring method (Csépes & Fekete, 2018b; Luoma, 2009). Since speaking is a productive skill, its evaluation is mostly subjective compared to receptive ones. Therefore, using rating scales

makes scoring and evaluation more reliable because not only task fulfillment but also the performance in terms of pronunciation, interactive communication, grammar and vocabulary are important issues for assessment (Csépes & Fekete, 2018b). The tasks used in oral production assessment can be either open-ended or structured depending on the level and ability of learners as well as the nature of the task. Description tasks, narrative tasks, instruction tasks, comparing and contrasting tasks, explaining and predicting tasks, decision tasks, role plays, simulations, reacting in situation tasks, structured tasks, interview, discussion, collaborative tasks, problem-solving task, short talk, group discussion, repeating, responding to prompts, and giving presentations are commonly used oral production tasks to assess speaking performance of language learners (Csépes & Fekete, 2018b; Heaton, 2011; Luoma, 2009; Spratt, Pulverness & Williams, 2016).

2.2.5.6. Assessing writing

Writing is regarded as ‘productive skill, social and cultural phenomenon, and cognitive activity’ (Weigle, 2009, p. 14). It is a *productive skill* and the performance of language learners can be directly observed through imitative, intensive (controlled), responsive, extensive writing tasks (Brown, 2004). Hamp-Lyons (2018) has drawn attention to the fact that writing in a second or foreign language is much more difficult than writing in one’s native language; so, assessing writing ability in foreign language is too important to monitor language development as well as to motivate learners to write. It is also important to choose appropriate writing tasks for language learners while assessing. Furthermore, while developing writing tasks, there are some issues to be considered. For example, test purpose, language use, language test performance, writing as performance assessment, test usefulness, genre, time, instructions, choice of tasks, transcription mode (computer or hand), use of dictionary, content/topic, tone/style, linguistic choices, rhetorical choices and scoring method need to be carefully taken into account (Hamp-Lyons, 2018; Weigle, 2009). The scoring method is mostly subjective and may be holistic, but it can be more reliable and analytic when rating scales are used to evaluate writing performance. In assessing writing skill, letters, reports, memos, messages, instructions, accounts of past events, composition, mechanics as punctuation and spelling, style and register, controlled writing, portfolio, copying, jumbled words, labelling, form filling, completion, story writing, and picture description are commonly

used techniques and tasks (Heaton, 2011; Spratt, Pulverness & Williams, 2016; Weigle, 2009).

2.2.5.7. *Assessing young language learners*

Young language learners are ‘those who are learning a foreign or second language and who are doing so during the first six or seven years of formal schooling [...] children who are in primary or elementary school [...] the ages of approximately five and twelve (McKay, 2008, p. 1). Since their social and cognitive processes differ from adult learners, language teachers need to have some kind of knowledge about their mental and social properties so that they can assess their language ability appropriately because assessment may have positive or negative impact on young learners’ language learning (McKay, 2008). Therefore, assessment is an important factor to motivate young learners in learning foreign languages. Language teachers can assess indirectly classroom language use of young learners or directly assess their language ability through tasks. In terms of classroom use, language teachers can make use of incidental observation, planned observation, on-the-run assessment, conferences, portfolios, contracts of work and project, self/peer assessment, class tests/tasks, keeping records, and teacher’s progress folders for the purposes of monitoring language development process of young learners (McKay, 2008). When it comes to assessing young learners’ language ability, language teachers generally assess their oral language in which grammar, vocabulary and pronunciation are also evaluated altogether (i.e., interaction, news telling, storytelling, etc.), reading (i.e., oral reading, miscue analysis, picture matching, etc.) and writing skills (i.e., speech bubbles, completing a story, etc.) (McKay, 2008). Their listening skills are also assessed through such tasks as placing objects, identifying people, multiple-choice pictures, color and draw (Hughes, 2003). To sum, while assessing young learners, not only the age but also the characteristics as well as the purpose of language learning need to be considered.

2.2.6. Importance of testing, assessment and evaluation in education

Language testing, assessment and evaluation are part of every education system in the world. They provide useful information to improve all the concerned issues such as teachers, students, learning, syllabus and the system itself in education. They are also used with different purposes such as selection, placement and determination of language

learning development to make progress in language education. For example, Bachman (2005) has stated that scores got from tests are very helpful to understand at which level language learners are, and accordingly, make some decisions to promote learning more and make them more successful in their language performance. Likewise, Green (2016) has indicated that assessment has some functions such as preparing language learners for their future exams (i.e., standardized-test), differentiating capable learners from less capable ones and enhancing more achievable learning, all of which serve useful information about learning/teaching the target language. Similarly, Hughes (2003) has noted that there is a need to test language learners to make comparisons about their language abilities and consequently, cultivating their learning more. Moreover, Brennan (2015) has emphasized that assessment can be used to serve many purposes: At the classroom level, it is helpful to improve both teachers' teaching and learners' learning by means of formative and summative assessment as well as students' reflections; on the focus of accountability, however, it is beneficial to enhance and regulate the group of learners (the class) as a whole, the effectiveness and quality of schools and system. In the same vein, Bachman and Palmer (1996), Heaton (2011) and Hidri (2018) have discussed that testing is important due to certain benefits such as providing information and evidence about language teaching/learning to make further decisions, supplying feedback about how much efficient the syllabus, curriculum, program and system are, and diagnosing students' weaknesses and strengths. From a different perspective, McKay (2008) has focused on assessment of young language learners such that assessing children's language development functions as providing evidences of their language performance and motivating them or demotivating them to learn more due to the positive and negative impact of assessment. In this issue, motivating learning, Heaton (2011) has argued tests can be used as reinforcing as well as motivating devices to motivate language learners to be more interested in and enthusiastic about language learning. Recently, although language learners have still cared much more about external exams rather than learning the target language itself, there has been a significant emphasis on formative assessment than summative one, which makes learning more motivating for learners (Gebril, Boraie & Arrigoni, 2016).

After all, testing, assessment and evaluation in education can be used with different purposes, and they are useful to understand whether teaching/learning the target language meets the needs of learners, expectations of teachers and objectives of the syllabus or

curriculum. Since there is a reciprocal relationship between teaching and assessment, the impact of assessment on teaching or the effect of teaching on assessment can be both positive and negative in line with the decisions made. Therefore, if assessment is expected to be useful, the impact of it on teaching should be taken into account while deciding and planning the next step according to the results of assessment.

2.3. Language Assessment Literacy (LAL)

The concept of literacy has been prevalent over the years in every field. Traditionally, it refers to the ability to read and write. Yet, because of the changes in society as well as information needs, its meaning has also expanded to include more critical and effective use of such ability for certain goals and areas (Taylor, 2013). Historically, the term of literacy was named firstly as alphabet-based literacy (just reading and writing skills) (1950s), secondly as functional literacy (knowledge put into practice) (1970s), thirdly as critical literacy (critical theory and pedagogy in education) (1980s), and lastly as new literacies (other areas such as information technologies, economy, society, business, health, law, etc.) (1990s-to date) (Ozarslan, 2017). The last category is very expanded and has still been developing. It covers a body of knowledge, skills and competence required where the term of literacy is used. For example, academic literacy, cultural literacy, computer literacy and emotional literacy reflect the ability to use associated domain knowledge and skills appropriately. Recently, people have needed to develop various literacies for specific sociocultural and functional dimensions to keep up with new improvements in the related field, and language assessment is just one of them (Taylor, 2013). In fact, the conceptualization of language assessment literacy has begun with the term of assessment literacy; so, the meaning of assessment literacy in general education has narrowed down to foreign language domain over the years due to the specific competencies required in foreign language education.

2.3.1. The concept of assessment literacy (AL)

The term of assessment literacy (AL) was coined firstly by Stiggins in 1991 within the context of general education. Basically, *assessment literacy* refers to the understanding of what assessment is, which to use, where and when to use, how to practice, how to benefit from its results, and why it is used and necessary (Stiggins, 1995). In other words, assessment literacy is ‘the understanding and appropriate use of

assessment practices along with the knowledge of the theoretical and philosophical underpinnings in the measurement of students' learning' (DeLuca & Klinger, 2010, pp. 420-421). Thus, the knowledge, theory and conceptual information are needed to be acquired along with necessary skills of assessment in order to perform effective assessment practices.

There have been many assessment and testing methods and tools used widespread across the world, especially, the standardized-tests, whose consequences have an impact on several people either positively or negatively in line with the quality of such methods and tools. Therefore, the knowledge of assessment concepts and procedures are very influential in educational decisions (Howerton, 2016; Popham, 2011). In the same vein, not only teachers and students but also other stakeholders such as principals and parents are affected by testing and assessment procedures. Thus, the meaning of assessment literacy has broadened and globalized to make all the concerned people in education to become aware of it (Dunkley, 2015). Referring to Stiggins, Green (2016) has argued that there are three levels of assessment literacy: 'functional level (public officials use assessment results); practical level (teachers and educators use and generate assessment data); advanced level (specialists in educational measurement generating data for others to use)' (p. 16). However, regardless of the roles of all the concerned people, each stakeholder should know what sound and unsound assessment is if they become assessment literate (Stiggins, 1995). In order to perform sound, namely, effective assessment, and to prevent negative impacts of unsound assessment, the concerned stakeholders are expected to possess certain features. Accordingly, Stiggins (1995) has put forward that assessment literacy requires to be able to identify clear purposes, focus on achievement targets, select appropriate tools, examine student success, and abstain from prejudice and disorder. Subsequent to Stiggins' arguments, there have been certain attempts to determine what kind of characteristics can represent the concept of assessment literacy. Mostly, such characteristics are called standards because those who have determined the content of assessment literacy have made it to be standardized for all the stakeholders around the world. Nonetheless, since teachers are acknowledged to be the most central stakeholders in education, the attempts to establish standards for teachers' assessment literacy have been substantial. The most famous one has been proposed by 'American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME & NEA)' in 1990. They have established

seven principles for teachers' competence in educational assessment including both pre-service and in-service teachers: '1) choosing appropriate assessment methods; 2) developing appropriate assessment methods; 3) administering, scoring, and interpreting the results of assessments; 4) using assessment results to make decisions; 5) developing valid grading procedures; 6) communicating assessment results; 7) recognizing unethical or illegal practices' (Mertler, 2003, pp. 9-10). So, since these standards in a way reflect the components of assessment literacy, both in-service and pre-service teachers should possess such characteristics in order to be assessment literate; otherwise, they cannot perform meaningful, effective and purposeful assessment in their educational life.

On the other hand, to have the knowledge and necessary skills to be assessment literate do not guarantee the quality of assessments because assessment practices might be affected by some factors. For example, Stiggins (1995) has indicated that fear of assessment, not enough time to assess, and public perspectives and attitudes towards assessment may set barriers to develop efficient assessments. In order to prevent such obstacles, providing professional development programs for in-service teachers, and giving courses to pre-service teachers at higher education in terms of assessment literacy training may be helpful (Stiggins, 1995).

As a conclusion, it can be said that assessment literacy is an important concept in educational assessment in order to provide better education. However, since every field necessitates its own knowledge domain, there must be different understandings and skills for each educational area. Therefore, language assessment literacy has been constructed in order to respond to the specific needs of language assessment.

2.3.2. The concept of language assessment literacy (LAL)

Language assessment is the integral part of teaching and learning. Therefore, foreign language teachers are expected to have both content pedagogical knowledge and assessment knowledge and skills in order to promote learning and regulate their teaching activities. However, in order to be assessment literate, foreign language teachers need to have a specific competence called *language assessment literacy (LAL)* which refers to the familiarity with and the ability to perform language-related assessment procedures.

In fact, the origins of LAL have dated back to the emergence of assessment literacy in general education by Stiggins in 1991, as mentioned before. Because of the recent growing emphasis on formative assessment and learning-oriented assessment as well as

the effect of sociocultural approaches on teaching/learning methodologies, LAL has emerged to respond to the needs of language teachers about assessment and evaluation, and also, to conceptualize the construct of LAL within foreign language education (Inbar-Lourie, 2017).

2.3.2.1. Definitions of LAL

In the literature, most researchers have defined the concept of LAL from the perspective of language teachers because teachers are considered as the central figures in classroom assessment (Giraldo, 2018). Therefore, generally, LAL is associated with other synonymous terms such as teacher assessment literacy, language teacher assessment literacy, teacher-based language assessment literacy, and classroom (teacher) assessment literacy. However, recently, there have been significant attempts to define LAL by including other stakeholders in the conceptualization (i.e., Taylor, 2013).

To start with, the concept of LAL refers to the knowledge, skills, and principles required for classroom assessment-related procedures like developing, administering, using, interpreting, analyzing and evaluating language tests and assessment tools as well as the awareness of assessment in a wider understanding within the scope of foreign language education (Lam, 2014; Pill & Harding, 2013). LAL also means language teachers' understanding of what and when to use assessment to obtain information about students' language abilities and progress and to give feedback in line with the results (Jeong, 2013). From the point of classroom assessment, H. Xu (2017) has defined LAL as EFL teachers' conceptions and skills regarding classroom assessment, which is put into practice in a planned or spontaneous way.

Apart from these, by doing research about textbooks used in language testing courses, Davies (2008) has identified that LAL refers to a three-level competence: Knowledge (the background in language description, context, and measurement), skills (the training in assessment methodology such as item writing and statistics), and principles (use of proper tests, fairness, ethics, and impact). Furthermore, Vogt and Tzagari (2014) have discussed that LAL is 'the ability to design, develop, and critically evaluate tests and other assessment procedures, as well as the ability to monitor, evaluate, grade and score assessments on the basis of theoretical knowledge' (p. 377). So, LAL is putting assessment theory and knowledge into practice.

However, Inbar-Lourie (2008; 2017) has emphasized the conceptualization of *LAL-language-focused* referring to language specific competencies by differentiating LAL from general assessment literacy. She has defined LAL as ‘to have the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results’ (Inbar-Lourie, 2008, p. 389). In addition, she has drawn attention to three rationales behind LAL as what, how and why. Based on these rationales, LAL refers to “the understanding the ‘what’ and performing the ‘how’ necessitates appreciation of the background and reasoning behind the actions taken, that is, the ‘why’” (Inbar-Lourie, 2008, p. 390). Hence, the most important feature in this definition is ‘what’ because it represents the language content which is assessed as a foreign language learning aspect.

Similarly, Giraldo Aristizabal (2018) has underlined language-related issues in defining the concept of LAL since foreign language teachers differ from other teaching subjects in terms of language components. LAL means the appropriate and fair use of assessment to develop students’ language learning on the basis of language assessment knowledge and skills both in classroom-contexts and in large-scale testing conditions. Thus, he has underscored the fact that language teachers need to have the knowledge about not only classroom-based assessment but also external testing procedures to help their students develop their target language. He has also indicated that most of teachers’ assessment practices have been shaped by their beliefs and prior knowledge of assessment in developing their own LAL.

Likewise, Stabler-Havener (2018) has argued that the competence of LAL requires the use of assessment knowledge and skills appropriately and ethically according to teachers’ own contextual characteristics. That is, each language teacher has own instructional context; so, they are expected to have the ability to use their assessment knowledge and skills in accordance with their students’ language needs such as proficiency to improve their learning, which show their competence of LAL.

On the other hand, although teachers are the central stakeholders in classroom assessment, LAL has gained new meanings in terms of the roles and effects of other stakeholders because LAL is acknowledged to be a ‘complex social practice’ among other stakeholders such as principals and test writers (Baker & Riches, 2017, p. 3). In this issue, Taylor (2013) has discussed the meaning of LAL from the perspective of the stakeholders involved in the language assessment procedures. In other words, though LAL refers to

understanding principles, applying theories in assessment practices, interpreting test data to give feedback in order to positively affect language learning in general (Taylor, 2009), there is a need to describe other stakeholders' roles in this process due to the growing numbers of people involved in testing and assessment recently. Therefore, she has conceptualized LAL in terms of three layers: Core, intermediary, and peripheral (Taylor, 2013). In the core, there are test makers and researchers. In the intermediary layer, there are course instructors, and language teachers. In the periphery, there are policy makers and general public. All of these layers reflect to what extent language assessment is understood, used, and affected by those involved. Hence, there are different roles, purposes or degrees of the stakeholders in language assessment practices according to their educational context. However, for classroom language teachers, LAL means to have the knowledge of language pedagogy in the first place, and then, to have technical skills about assessment as well as to apply and use assessment in correspondence to their beliefs, attitudes, local practices and sociocultural values.

Unlike the definitions discussed so far, Fulcher (2012) has provided an expanded definition for LAL by studying with EFL teachers. He focused on the sides of theory, practice and context about language assessment. He has defined LAL as:

‘The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals.’ (Fulcher, 2012, p.125)

This definition has been recognized as the most comprehensive one to date in LAL literature although it does not explicitly refer to foreign language components (Stabler-Havener, 2018). In contrast, it has reflected the characteristics that can be applied both to classroom and large-scale assessments in any EFL context (Giraldo, 2018) because the definition was constructed with help of in-service EFL teachers' conceptions and practices in Fulcher's (2012) study.

To sum up, all the definitions of LAL mentioned so far have common properties such as the understanding, familiarity, use, knowledge, skills, principles, and abilities about language assessment accompanied by the content pedagogical knowledge of foreign language teaching/learning. However, in the present study, Fulcher's (2012)

definition of LAL is prioritized because of its comprehensive dimensions emphasizing language assessment in EFL contexts. Moreover, since the current study is carried out in an EFL context with English language pre-service and novice teachers as well as course lecturers of language assessment, the term *Foreign Language Assessment Literacy (FLAL)* is preferred to be used in order to represent the characteristics of EFL context in which English is taught and learned as a foreign language; so, it is assessed as a foreign language.

2.3.2.2. Dimensions or components of LAL

Language assessment literacy has been extensively studied in recent times in order to conceptualize and establish its dimensions or components. When LAL research conducted up to now is considered, Stabler-Havener (2018) has discussed that there are four mainstreams regarding LAL models. The first one is about the standards of language assessment literacy, and its aspects of what-how-why; the second one is skills, knowledge, principles model; the third one is practices, principles, contexts model; and the last one is LAL stakeholder model. All of these models reflect the meanings and dimensions of LAL, especially from the perspective of teachers' competencies.

In the first place, Davies (2008) has asserted that LAL has three components: *Knowledge*, *skills*, and *principles*. The component of *knowledge* refers to the framework used in measurement, how language is defined, and how the setting is described in terms of language assessment. The component of *skills* is about the application side of language assessment; developing test, writing items, administering test, and analyzing the results. The component of *principles* includes how tests can be used properly and appropriately as well as ethically.

Secondly, Inbar-Lourie (2008) has argued that LAL is a multi-layered concept which responds to the questions of *what*, *how*, and *why* with respect to language assessment. '*What*' refers to the description of the language components to be measured; '*how*' is about the implementation side of assessment or the process itself; '*why*' is related to the rationale or reasoning of language assessment practices. Therefore, language teachers plan their assessment according to '*what*', choose appropriate methods and tools according to '*how*', and base their decisions and practices on '*why*'. In terms of this model of LAL, '*what*' is the most distinctive component because it is directly related to the foreign language content to be measured (Inbar-Lourie, 2008).

Thirdly, Scarino (2013) has divided the dimensions of LAL into two categories: *knowledge* and *process* base. *Knowledge base* is about the theoretical underpinnings of language domains which help to develop and use assessment in practice. *Process base* is the developmental phase of language assessment while examining and criticizing the conceptions of assessment. Then, both knowledge and process base are required as the main constituents to conceptualize LAL.

Fourthly, Fulcher (2012) has developed a model of LAL as result of his scale development with EFL teachers about language assessment. According to his model, LAL has three competencies: *Practices*, *principles*, *contexts* (see Figure 2.2). The dimension of *practices* refers to designing, developing, administering, evaluating both standardized and classroom tests as well as the familiarity with assessment procedures. The *principles* dimension is about the awareness of principles and ethical issues and concerns about language assessment in order to guide the practices. The dimension of *contexts* involves the underlying frameworks such history and society, the origins and rationales, and the impact of language assessment in a broader sense. However, such dimensions of LAL may not be valid for all the stakeholders because Fulcher (2012) has noted that that model was constructed with EFL teachers; so, it represents their LAL competencies.

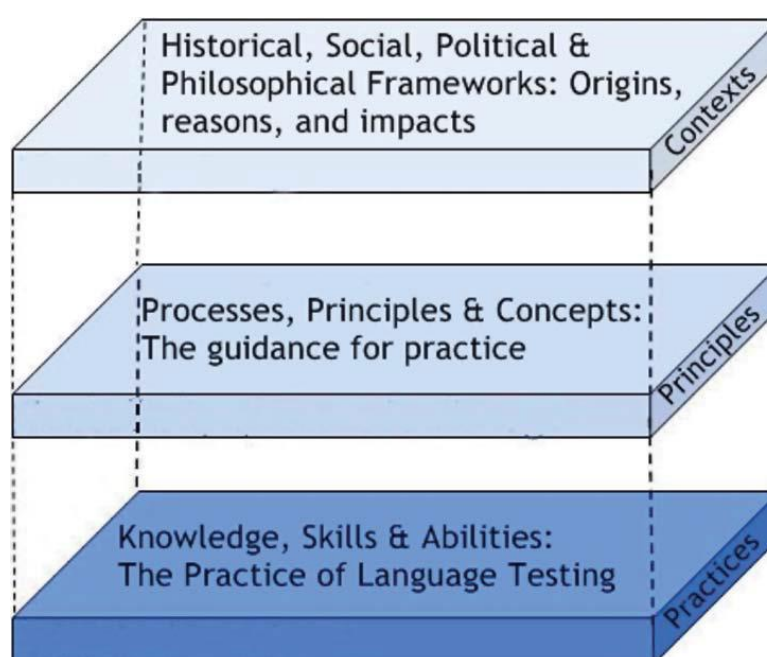


Figure 2.2. Language assessment literacy: An expanded definition (by courtesy of Fulcher, 2012, p. 126)

Fifthly, Hill (2017) has focused on only classroom-based teacher assessment literacy, and correspondingly, has developed a framework called *Teacher Assessment Literacy (TAL)* within the scope of language assessment. She has taken into account the classroom assessment practices such as what teachers do and look for, what sort of theories they use, what they believe, how their learners view assessment, and how the context influences the practices of assessment while constructing her framework. By answering such questions, she has proposed a three-dimensional TAL model for language assessment: *Practice, concepts, context*. The dimension of *practice* is about the planning, conducting and using assessments as well as possessing necessary knowledge and skills about language assessment in relation to the characteristics of students and qualities of assessment. The dimension of *concepts* is about teachers' own beliefs and understandings which shape their assessment procedures. The dimension of *context* is about the impact of students' views, the decisions of schools or supervisors about language assessment, which has an impact on teachers' classroom assessment practices. Thus, all three dimensions have shaped the constitutes of LAL at classroom teacher level.

Lastly, apart from teacher-based LAL models, Taylor (2013) has concentrated on other stakeholders who are also involved language assessment processes such as test writers and administrators. According to her conceptualization of LAL, each stakeholder has different competencies with different degrees of significance in constructing their LAL. There are basically four stakeholders in assessment process: Test writers, classroom teachers, administrators, and professional testers. However, their roles in assessment have been shaped according to three layers as core, intermediary, and peripheral, as discussed previously (see Figure 2.3). Language teachers are within the level of intermediary, which indicates that they have somewhat different constitutes of assessment literacy compared to other stakeholders.

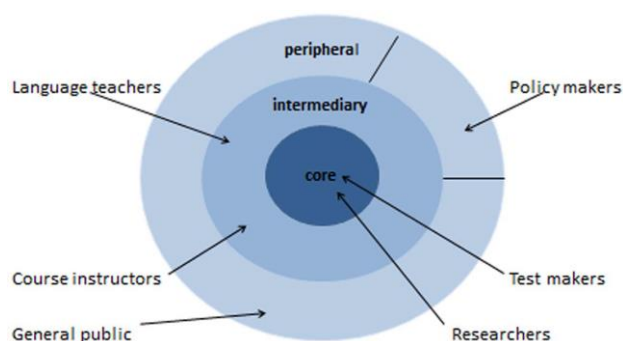


Figure 2.3. Levels of AL/LAL differentiated according to stakeholder constituency (by courtesy of Taylor, 2013, p. 409)

Thereupon, to illustrate the degrees of significance and to show the importance of different competencies according to stakeholders, Taylor (2013) has employed spider-web like figures (see Figure 2.4). In these figures, it is apparent that some competencies are more important for certain stakeholders compared to others. For instance, firstly, language teachers need to have the competence in the area of language pedagogy, and then, they must have technical skills of assessment. Their beliefs, attitudes, local contexts and sociocultural values have also an impact on their LAL profile.

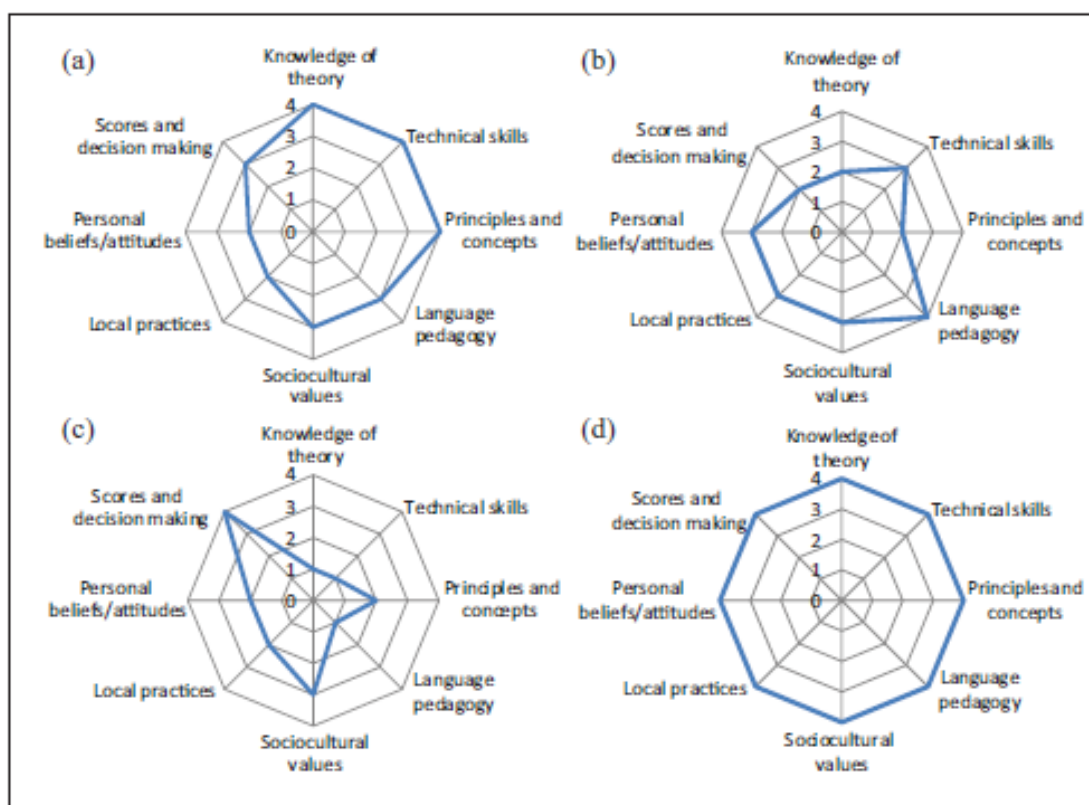


Figure 2.4. *Differential AL/LAL profiles for four stakeholders (by courtesy of Taylor, 2013, p. 410)*

As a conclusion, nearly all the dimensions or components of LAL discussed so far have illustrated mostly a three-layered model. The most important features of LAL, then, consist of knowledge, skills, abilities, principles, practices, concepts, origins, contexts and reasoning about language assessment.

2.3.2.3. *Characteristics of LAL*

Language assessment literacy is an important concept in foreign language education, and language teachers must possess this kind of competence in order to

promote language learning. Thus, there are some characteristics that indicate a teacher is assessment literate. For example, Fulcher (2012) has pointed out that to have a balanced proportion of knowledge, skills and principles about language assessment enables teachers to understand and use assessment practices properly.

Moreover, Stabler-Havener (2018) has argued that language assessment literate teachers must firstly define language proficiency, and then, possess related knowledge and skills of assessment theories and practices. Also, they must be able to execute assessment procedures appropriately and communicate their results according to their instructional context.

In addition, NTF (2016) has stated that teachers become assessment literate when they master in terms of sound assessment principles as well as the application of assessment methods suitable to their contextual needs. Though NTF (2016) has described the LAL characteristics of each stakeholder by their roles, from the point of classroom teachers, it has underlined that assessment literate teachers know the interaction between their instruction and assessment, have the ability to define their assessment objectives, have the ability to design and administer appropriate language tests and assessments, are aware of various assessment techniques available for use, have the ability to analyze and interpret the results to reinforce and motivate learning/teaching, know how to communicate with students, colleagues, parents and society by means of assessment findings, and are aware of LAL as a developmental feature which lasts throughout their teaching life.

Furthermore, most of the researchers have described similar characteristics of language assessment literate teachers. They have highlighted that assessment literate teachers understand what sound assessment is; are able to decide and apply assessment objectives in line with learning outcomes; know how to practice appropriate assessment methods; are able to choose or adapt suitable assessment techniques for specific purposes in specific contexts; interpret the results or the performance to regulate their own teaching as well as to improve students' learning and increase motivation; know how to report the results to others and accordingly, give feedback; have the ability to integrate the knowledge of assessment in the decision-making process of their instructional practices; understand the outcomes and expectations of society and policy about assessment; master in language specific competencies in relation to their assessment ways; are aware of several assessment tools to use; have the understanding of both large-scale and classroom

assessment; and are able to perform assessment ethically (Berry & O'Sullivan, 2016; Gotch & French, 2014; Huang & He, 2016; Inbar-Lourie, 2013b; Khadijeh & Amir, 2015; Newsfield, 2006; Rogier, 2014; Taylor, 2012; Zolfaghari & Ahmadi, 2016).

Besides, Looney, Cumming, van Der Kleij and Harris (2017) have proposed a new term to describe the features of *teacher assessment literacy (TAI)* as result of their research. They found that TAI consists of the dimensions of 'I know, my role, I am confident, I feel, I believe' regarding assessment literacy. Therefore, they concluded that TAI reflects the dimensions of knowledge, skills, practices, confidence, beliefs, frameworks, purposes, uses and implementations of assessment, all of which are the sings of assessment literate teachers.

Similarly, Ukrayinska (2018) carried out a study to find out the characteristics of *classroom language assessment literacy skills*. She focused on EFL classroom teachers in the Ukrainian context. As a result of her study, she described the characteristics of assessment literate teachers as being able: 'to plan assessments, to design assessments, to administer assessments, to interpret assessments results, to apply the results of assessments accurately and effectively, to discuss the results of assessments with parents and students, to evaluate classroom-based tests, and to identify the uses and misuses of externally mandated large-scale assessments' (pp. 363-364).

Likewise, Giraldo (2018) studied on Davies' (2008) dimensions of LAL as knowledge, skills, and principles in line with the characteristics of language assessment in order to expand the competencies. Thus, he presented a comprehensive list for LAL dimensions, and this list is made up of 66 subskills about LAL. In terms of *knowledge*, awareness of applied linguistics, theory and concepts, and own language assessment context are important features. With regard to *skills*, instructional skills, designing language assessment, skills about educational measurement and technological skills are the parts of LAL skills. Pertaining to principles, awareness and critical perspectives of language assessment are significant indicators of LAL aspects. As a conclusion, he emphasized that all of the three-dimensional model of LAL accompanied by 66 subskills represent the descriptors of language assessment literate teachers.

As a consequence, in a general sense, assessment literate teachers know what language assessment means, what kinds of theories and concepts there are behind assessment, are aware of assessment methods and uses, know how to assess and are able to implement methods of assessment appropriately in line with specific purposes, needs

and contexts, and know how to use assessment results to improve both their teaching and students' learning.

Considering the Turkish context in this sense, since most of English language teacher candidates are appointed by Ministry of National Education (MoNE), there are some expected qualifications from the graduates to possess. Such qualifications are called ELT competencies and can be regarded as the standards for Turkish EFL teachers. MoNE-ELT Competencies (2008) are made up of five dimensions: Planning and organizing the processes of English language teaching/learning; developing students' language skills; monitoring and assessing language development/progress; collaboration/cooperating with school, family, and society; and maintaining professional development in the field of English language teaching. Out of such competencies, monitoring and assessing language development/progress is very important because Turkey is a test-driven country emphasizing high-stakes examinations, and also, Turkish context is an EFL context which requires students to be assessed regularly in terms of their language development as they do not use language in their daily lives much. In terms of the qualification of monitoring and assessing language progress, it has certain characteristics which reflect EFL teachers' assessment literacy such as the ability to determine the aims of measurement and evaluation, to use instruments and methods of measurement and evaluation, to interpret the results of assessment and give feedback to identify the language development, and to use the results of assessment in teaching practices (MoNE-ELT Competencies, 2008). All such characteristics can be employed as the guidelines to represent the characteristics of language assessment literate teachers in the Turkish context.

2.3.3. Importance of LAL in foreign language education

It is widely known that assessment is an inseparable element of teaching because it helps learners to develop their foreign language knowledge and skills. Recently, foreign language teachers have to be competent educators because their roles have changed due to more emphasis on the notion of formative assessment to improve learning rather than only to measure learning as the product (Csépes, 2014). They have become both teachers and assessors of their own teaching contexts. Hence, they must have content-pedagogical knowledge about foreign language teaching as well as related competencies about language assessment. Then, language assessment is an important issue in foreign

language education. It is because assessment influences the instruction of teachers, provides information about how much students learn the target language and whether they are able to use it, and what needs to be done to promote their learning more successfully (Rogier, 2014). In addition, good practices of language assessment are important since it is useful to give information about the content of teaching, helpful for teachers to monitor students' development, beneficial for students to see their own learning progress, and advantageous for preparing students for future exams and assessment throughout their educational lives (Herrera & Macias, 2015).

To be able to perform effective language assessment procedures, teachers need to have some kind of knowledge and appropriate skills to maximize students' learning by regulating their teaching practices. Therefore, they must be assessment literate to respond to the needs of their instructional contexts. Then, LAL is a vital competence for foreign language teachers. There are many reasons or factors regarding the importance of LAL discussed in the literature. For example, Newsfield (2006) has argued that if teachers know assessment terminology, procedures to carry out, interpretation ways of results, and assessment ethics, then, these characteristics as the constituents of LAL will allow them critically to evaluate their classroom language progress to enhance learning in a more achievable way while motivating learning the target language. Furthermore, to be assessment literate is fruitful in helping teachers perform their assessment methods according to their own instruction, the purposes of teaching, and the needs of students while building a relationship between teaching and assessment for higher achievements of students in the target language (Bayat & Rezaei, 2015). LAL is also helpful in choosing best alternatives for assessing learning, monitoring the progress, and performing better teaching practices (Zolfaghari & Ahmadi, 2016). Besides, to have a higher level of LAL is useful to develop professionally while maximizing and motivating foreign language learning (Buyukkarci, 2016). However, not only teachers but also other stakeholders need to be assessment literate to help the process of assessment to be performed in an effective way to promote learning as well as to eliminate unsound assessment practices to prevent negative washback on learning and teaching (Djoub, 2017; Zolfaghari & Ahmadi, 2016).

To conclude, as White (2009) has indicated that improved LAL is helpful to undertake better assessment methods, which makes both teaching and learning to be developed and motivated for better foreign language learning outcomes.

2.3.4. Review of relevant research studies on LAL

Language assessment literacy has been an important issue in foreign language education as well as foreign language teacher education. It is because this concept is regarded as the indicator of a good deal of knowledge and effective practices and skills in terms of language assessment.

Research into the relationship between assessment and teaching can be considered to lead LAL research to expand because there have been changes in the notions of foreign language teaching/learning which have an impact on language assessment procedures, and reciprocally, methods of assessment have influenced teaching practices (Clapham, 2000). Therefore, historically, the impact of these two concepts have been studied from different perspectives. Firstly, the washback effect of assessment classroom teaching has been studied extensively; secondly, teachers' roles in assessment and teaching, the phases of designing assessment practices, authenticity and appropriacy of tests have been investigated; thirdly, learner attainment through the effect of language programs and their success have been explored; lastly, the role of teachers as assessors and classroom assessment accompanied by such topics as purposes, criteria, cultural context, and levels of assessment have been searched (Rea-Dickins, 2004). Therefore, recently, there has been an emphasis on classroom assessment and the role of teachers as assessors, which makes LAL to be a prominent concept. It is because rather than adopting testing as a traditional way of evaluation method, teachers are expected to create their own assessments according to their own contextual needs in order to improve and monitor language progress of learners (Tsayari & Vogt, 2017). Thus, teachers need to be assessment literate to respond to such needs in their own contexts while teaching foreign languages. As a conclusion, it should be noted that there have been more studies with in-service EFL teachers than pre-service teachers in terms of LAL.

2.3.3.1. Research studies conducted in in-service EFL contexts

In foreign language education settings, research into the LAL of in-service teachers have been studied from different points of views. But mostly, the effect of demographic variables and teachers' beliefs on LAL, the assessment practices and the needs of teachers about LAL, and the impact of workshops designed as in-service training on LAL have been the notable topics on the concept of LAL in recent years.

Pertaining to demographic variables, Y. Xu and Brown (2017) studied on the validation of 'Teacher Assessment Literacy Questionnaire (TALQ)' through examining the effects of some demographic variables of EFL teachers working at university on their assessment performance in China. The researchers also investigated the level of their assessment literacy. They employed an adapted version of TALQ, which was paper-based questionnaire and included 24 scenarios about language assessment and 7 demographic variables such as age, qualification, experience and training. A total of 891 EFL teachers from 7 regions of China responded to the questionnaire. The results yielded that the adapted questionnaire needed revising to reflect Chinese university contexts. The assessment literacy levels of the sample were also found at the basic level in certain domains, and thus, needed to be improved. However, all the demographic variables were found to have small effects on their LAL performance. Hence, the researchers concluded that though there was little effect of demographic variables, assessment literacy of EFL teachers needed to be promoted, specifically, in the areas of policy and professional standards.

Likewise, Fard and Tabatabaei (2018) were interested in the levels of assessment literacy of EFL teachers in Iran. The aim of their research was to determine to what extent EFL teachers were knowledgeable about language assessment. To analyze this issue, they administered 'Classroom Assessment Literacy Inventory' to 52 EFL teachers working at both public schools and private institutions. Their responses to the questionnaire showed that their assessment literacy levels were too low and needed to be enhanced. The researchers discussed that the participants were not aware of what sound assessment means and did not feel ready to handle the problems in classroom assessment.

With the aim of exploring LAL from the perspective of new literacy studies, Duboc (2009) investigated 3 elementary school EFL teachers in Brazil. She based her study on interpretive qualitative study and explored the conceptions and practices of the participants who taught English to 6th grades. The interviews, observations and written tests were used to collect data from EFL teachers. The findings revealed three main problems: The first was that evaluation was perceived as the same with measurement. The second was the participants relied on linguistic code assessment, namely, grammar and vocabulary, too much, and thus, they neglected other areas of language development. The last was that teachers employed mostly written tests rather than classroom or alternative assessment. Therefore, the need to redesign the curriculum at teacher

education programs was reported for better LAL training to prepare future teachers for their assessment conceptions and practices.

Another study dealing with beliefs and practices of EFL teachers about language assessment was conducted by Wach (2012). She investigated the beliefs and roles of both school teachers and university teachers about language assessment in terms of being native or non-native speaker of English. Therefore, the school type and the feature of being native or not in English language were explored whether they affected the beliefs and roles of teachers in Poland. To address this issue, 87 participants (42 of them were university teachers, and 45 of them were school teachers) took part in the study. Out of 42 university EFL teachers, 12 of them were native speakers of English while 30 of them were non-native speakers. As for school EFL teachers, out of 45, there were 17 primary, 20 middle and 8 secondary school teachers in the research. The participants had teaching experience ranging from 6 months to 30 years. By using open-ended questionnaires, the researcher found that university teachers were more aware of assessment functions, areas and features than school teachers. Nearly all the participants focused on grammar and vocabulary assessment, and they considered assessment is separate from teaching. In terms of being native or not, it did not differ much regarding the opinions of teachers. As a result, the researchers suggested that EFL teachers needed to be trained in terms of different assessment types valuing the evaluation of different language areas and skills.

Munoz, Palacio and Escobar (2012) carried out their study in order to examine the beliefs, perceptions and practices of EFL teachers about language assessment. They studied with 62 EFL teachers who were working at the language center of a private university in Colombia. To explore their such issues, surveys, written reports and interviews were collected. The analysis yielded that there was a mismatch between their beliefs and practices; that is, they believed reflection, self-assessment and formative assessment are good and should be used whereas their assessment practices concentrated on more summative evaluation of their students. Thus, the researchers recommended professional development programs may help teachers practice what they believe rather than sticking only to summative assessment.

From the perspective of ELT assessment literacy ideology, Hakim (2015) investigated the awareness levels as well as the ways of assessment practices of EFL teachers. In other words, the researcher examined the relationship between what teachers believed and what they practiced in their classrooms about language assessment. 30

female instructors who had different nationalities, qualifications and teaching experiences, and were working at the English Language Institute at university in Saudi Arabia responded to the researcher-developed questionnaire about assessment. The results showed that although EFL teachers had adequate knowledge of assessment tools, their practices did not develop much. However, those who were more experienced in terms of teaching and assessing had a clearer understanding of what assessment means, what kind of components assessment have, and how to implement a testing plan. Thus, teaching experience had effect in the perceptions of EFL teachers' assessment literacy.

Similarly, Jannati (2015) explored the perceptions and practices of EFL teachers about their language assessment in line with their teaching experience. Therefore, the researcher organized three groups of EFL teachers according to their teaching experience as being high, mid and low-experienced teachers. Conducting semi-structured interviews with 18 EFL instructors working at language institutes in Iran, the researcher found that nearly all the instructors had a good amount of knowledge about basic concepts and terms regarding language assessment. However, their knowledge did not reflect on their practices; that is, they could not put their knowledge into practice. With respect to experience, it did not affect EFL teachers' assessment practices. Thus, the instructors needed to learn how to practice their assessment knowledge in their classrooms.

From the point of different teaching subjects, Zolfaghari and Ahmadi (2016) examined what teachers from different academic backgrounds understand about assessment literacy, and whether the components of assessment literacy differ among such backgrounds. 32 high school teachers, all of whom were native speakers of Persian, took part in the study. Their subjects were humanities, science, English language, and non-English language (Arabic and Persian languages) teachers. The researchers conducted a semi-structured interview including demographic information and assessment literacy questions. The audio-recorded data were thematically analyzed, and the findings pointed out that the participant teachers mentioned only two assessment methods as continuous/diagnosis and continuous/final. Most of them were familiar with certain assessment methods (i.e., peer-assessment) and practiced them, but unconsciously. They thought that to be qualified in assessment literacy was affected by both teachers' and learners' attributes. In terms of literacy components, many participants agreed that assessment literacy is a mixture of content, ethical, emotional and assessment concepts, but they believed since the content of each subject necessitates different

competencies, assessment literacy differs in this point. The participants suggested training programs about assessment should be revised.

Apart from these, Hudaya (2017) designed a study to investigate whether EFL teachers felt prepared to assess, what kind of principles they applied in assessment, and what sort of problems they faced in assessing. Hence, the researcher developed a questionnaire which consisted of both closed-ended and open-ended questions about language assessment and principles. 43 in-service EFL teachers from different teaching stages (junior, elementary, senior, university and kindergarten) in Indonesia responded to the questions. The results produced that more than half of the participants felt prepared to assess their students' performance. Also, most of them agreed that practicality, reliability, validity, authenticity and washback were the important principles in assessment practices. Though many teachers were found to be assessment literate, they reported that they had sometimes difficulties in giving feedback.

Djoub (2017) conducted a worldwide web-based questionnaire in order to scrutinize whether assessment literacy had an impact on the views and practices of language assessment. 45 teachers from different countries and school types, and also, from different language contexts (both ESL and EFL) answered the questions about teaching experience, assessment training, the components of sound assessment and the role of institutions about language assessment. Besides, 10 of them were interviewed about those issues. The findings indicted that language assessment training did not assure being assessment literate. Moreover, many teachers emphasized that as long as the content, the tools used and the way administered were decided and applied, then, such assessment could be called as sound assessment. Most of the participants also highlighted they needed more autonomy, training, collaboration, feedback and supplementary materials regarding language assessment practices.

Carrying out an action research, Giraldo Aristizabal (2018) investigated the beliefs and practices of EFL instructors related to language assessment literacy while designing an achievement test. 60 EFL instructors working at a language institute in Colombian University participated in the study. They responded to questionnaires and were interviewed. Also, their testing and assessment documents were analyzed. The findings revealed that according to the participant EFL instructors, a test should consist of four principles as reliability, validity, authenticity and washback, all of which could leave positive impact on teaching and learning. Even though they believed so, they were not

able to reflect those principles in their practices successfully. As a whole, the experience of designing an achievement test was found beneficial to their professional development.

Other than this kind of research about the beliefs of language assessment literacy, some researchers focused on in-service teachers' practices of language assessment. For example, Y. Xu (2016) explored how teachers planned their assessments. Therefore, she contacted with 20 EFL teachers working at contest-winning university in China and collected their lesson plans. She also conducted interviews with them about their assessment practices. The qualitative analyses showed that the lesson plans involved assessment-related topics such as objectives, assessment tasks and student backgrounds. While the plans stuck to the learning goals, the rubrics or checklists did not exist to evaluate their students' performance, and also there were some mismatches between objectives and assessment tasks. Lastly, nearly all the teachers underlined that they planned their assessments intuitively and in line with their instructions.

In addition, Tsagari (2013) investigated what university students thought about assessment and what kind of assessment practices were used in their EFL classes in Cyprus. In other words, language testing and assessment practices performed in higher education were examined. 41 undergraduate students from different programs (primary education, business, science, secretarial studies, computer studies, graphic) responded to a questionnaire about the assessment types conducted in their EFL classes. The most common testing type was written tests whereas portfolio was not used much. Their assessment was based on the purpose of measuring the language ability; however, listening and speaking tests were not administered. In grammar and reading tests, multiple choice was preferred the most, and in writing exams, essay writing was used the most. The researcher emphasized that assessment practices in EFL classes in higher education tended to be summative rather than formative; and hence, they did not improve learning, just measuring what students knew.

Similarly, Klinger (2016) designed a study to examine how EFL teachers working as professors at university conceived assessment and how assessment was performed at their university. Based on a qualitative study as phenomenology, the researcher tried to discover the understandings, beliefs and practices of EFL professors' assessment practices in Colombia. 4 full-time professors from 'Linguistics and Languages Department' and having more than 15 years of teaching experience about LAL training took part in the study. Interviews and documents were employed to collect the data from

the sample. The analysis of the data showed that assessment was mostly used to measure students' knowledge. Though professors appreciated the importance of assessment for formative purposes, they admitted that they were not able to assess properly their students. Besides, students were not informed about how they would be assessed for the course. Therefore, further training for assessment was needed for the professors.

Furthermore, Berry, Sheehan and Munro (2017) analyzed the actual assessment practices of EFL teachers in line with their attitudes and needs about language assessment. They designed a three-phase study to address those issues. In the first phase, 3 experienced international EFL teachers working at a British university were interviewed. In the second phase, those 3 EFL teachers were observed. In the last phase, 48 experienced EFL teachers were selected for focus group discussions. The findings revealed four issues about language assessment. Many participant teachers underlined the fact that they lacked training on language assessment. They also mentioned testing rather than assessment, and stated they needed materials for assessing language skills. In general, teachers were not familiar with the concept of language assessment literacy.

Performing a three-year longitudinal study with the novice EFL teachers in China, H. Xu (2017) tracked their development of language assessment literacy in their first year of teaching. During that process, he executed 26 classroom observations and took fieldnotes about 4 EFL teachers, conducted 48 interviews with 2 EFL teachers, and collected journal entries of 171 teachers. After analyzing all the data, he found a three-phase development of LAL regarding novice EFL teachers: In the first stage, teachers began to acquire practical skills for the planned classroom assessment; in the second stage, they became more aware of the connection between teaching objectives and planned assessment practices; in the third stage, they started to perform more spontaneous assessment in their classes, especially in terms of formative assessment. Thus, he concluded that especially the improvised stage of performing assessment was the sign of language assessment literacy development, and novice EFL teachers improved over time regarding their assessment practices.

Beginning with a research project, Fulcher (2012) designed his study in order to develop a survey instrument which examines learning needs in assessment. He focused on the assessment training needs of in-service language teachers. All 278 participants from different countries and from different language teaching areas contributed to the study. The analysis yielded a four-factor survey: 'test design and development', 'large-

scale standardized testing', 'classroom testing and washback', and 'validity and reliability'. Accordingly, the researcher proposed an expanded definition of language assessment literacy which comprised of three components as contexts, principles, and practices. Hence, those components were found to reflect the training needs of language teachers in assessment and testing. The last finding was about the materials used in assessment training, and the sample found them insufficient to respond to their needs.

Differently from other studies, Razavipour (2013) concentrated on the issue of stakeholders; that is, how stakeholders viewed LAL, and what kind of needs the stakeholders mentioned. There were three types of stakeholders: LAL test developers, LAL instructors and LAL test-takers (54 MA TEFL students). Their perceptions, their needed areas of assessment training, and the problems related to assessment were investigated through surveys. The tests constructed by the test-developers were also analyzed. All the participants were the part of the assessment course doing their Master of Arts (MA) training in Iran. The gathered data were examined, and crucial findings were reached. While test-designers focused on knowledge and theory, the instructors and MA students tended to underline the skills of assessment. Hence, a gap between what test-designers believed and practiced and what teachers and students thought about language assessment was detected. The researchers concluded that to fill that gap, training courses both at the level of undergraduate and graduate program should be designed to improve LAL of all the stakeholders.

Conducting a comprehensive study across Europe, Vogt and Tsagari (2014) explored the training needs of in-service EFL teachers. They collected data from 853 participant EFL teachers from 7 European countries (Cyprus, Macedonia, Germany, Greece, Italy, Poland, Turkey) by means of a questionnaire about LAL training needs. They also conducted interviews with 63 EFL teachers from 3 countries (Germany, Greece, Cyprus). The findings revealed that teachers' LAL did not develop much, and few of them received training about language assessment. Besides, those who received training underscored the fact that training did not prepare them for their work. Also, teachers did not have much idea about the recent assessment procedures. Thus, the researchers indicated that the training about all the LAL features was needed responding to the needs and issues of local educational contexts of EFL teachers.

In their another study, Tsagari and Vogt (2017) focused on the qualitative part of their comprehensive study mentioned above and conducted in 2014. They explored the

perceptions and training needs of in-service EFL teachers about LAL. A total of 63 in-service teachers working at primary and secondary schools in Germany ($n=25$), Cyprus ($n=16$), and Greece ($n=22$) took part in the study. The researchers administered semi-structured interviews to examine the training biographies of the participant teachers. The results showed that nearly all in-service EFL teachers relied on pen-and-paper tests, and traditionally, assessed their students' especially grammar and vocabulary. Their marking procedures were only about correcting the errors and mistakes. Moreover, although teachers were familiar with alternative assessment methods, they did not benefit from them in their practices. In terms of their LAL training, teachers indicated they received little training, and learned most of the things by experiencing them in their classes. Still, they could not determine their localized training needs about language assessment. Thus, the researchers stated that teachers did not have adequate LAL because of the insufficient training, and teachers' conceptions and perceptions should be transformed into more formative assessment perspective to update them.

In the same vein, Yan, Zhang and Fan (2018) explored the training needs of EFL teachers through their development of LAL. They studied with 3 middle school EFL teachers from China and performed semi-structured retrospective interviews about how the contextual and experiential factors affected their practices of assessment. The findings indicated that teachers developed their LAL as a result of their intuitions and experiences rather than training because their training had focused on more assessment theory than practice. Therefore, they needed training in terms of practicing skills of assessment. Another finding was that both context and experience were highly influential in their LAL development as well as their assessment practices. Specifically, though they preferred and wanted to make formative assessments, the educational policies did not permit them to do such practices because of the country's emphasis on high-stakes entrance examinations.

Studying with ELT practitioners, Sheehan and Munro (2017) designed a project to find out what teachers needed to know and what kind of materials required for their assessment training. To explore this, three-stage research was executed. Firstly, 3 EFL teachers who had international teaching experiences and working at a British university at that time of the study was carried out were interviewed and observed. Secondly, 48 EFL teachers working at language teaching centers took part in focus group interviews. Lastly, 20 teachers participating a workshop in the conference of IATEFL gave their

opinions about language assessment. All the data analysis yielded that tests were very influential in classroom assessment practices, and teachers seemed to have lack of confidence in making assessments. Besides, teachers stated that they needed training in terms of specifically practical assessment because they developed their LAL through only experience. Lastly, they indicated that they did not want to spend time on training too much; thus, it is better to take examples of ready-made materials to be used in assessment.

Regarding the workshops designed as in-service training about LAL, Gebril, Boraei and Arrigoni (2016) prepared and implemented a training project for EFL educators in Egypt. Since Egypt is a text-driven country, there is a need to strengthen the knowledge and skills of assessment among EFL educators. In line with this need, a two-level training was organized; firstly, reflections on assessment practices were carried out and teachers were made to adapt testing materials into their contexts, and then, the collaboration among the practitioners about assessment was built via social media platforms. A total of 48 participants attended the training, and it lasted for 4 days including 16 sessions about language assessment. During that training, reflective journals, plans of assessment, post-training surveys and follow-up activities were collected to be analyzed. The findings revealed a positive development in teachers' LAL; however, the sample found the training was too intensive due to its little time for such a rich content. Thus, there would be a longer training project for EFL teachers to improve more about language assessment literacy.

Likewise, Baker and Riches (2017) designed a workshop as in-service training with the aim of monitoring the LAL development of EFL teachers as well as language assessment specialists. During a one-week workshop training, 120 secondary school EFL teachers from Haiti none of whom took assessment training before attended the training. In addition to EFL teachers, Canadian facilitators as the specialists took part in the workshop. In the training, the feedback given on examinations were collected, and surveys and interviews were conducted with regard to language assessment. At the end of the training, both groups of participants showed positive improvement. However, there were challenges in terms of collaborative decision-making in the group of EFL teachers. The specialists, on the other hand, were more knowledgeable than EFL teachers about assessment. Hence, the researchers concluded that there should be collaboration between teachers and specialists such that specialist may help teachers to enhance their assessment knowledge and skills more.

Based on a longitudinal study, DiBiase-Lubrano (2018) investigated the change of instructors' LAL after attending a series of workshops at university. Also, the effect of those workshops was examined. As a professional development opportunity, assessment workshops were offered over six semesters at Yale's Center for Language Study. In those workshops, 20 different topics of assessment were discussed. At the end of the workshops, the surveys about the content and format of the training as well as the suggestions for further workshops were administered. Only 16 lecturers responded to the online survey. As a whole, the workshops were found beneficial, and the participants gained much more knowledge about technical and logistical issues such as reliability and validity. The most useful workshop was using Excel program to analyze test items according to the sample. For future workshops, they suggested rating scale performance regarding speaking and writing skills would be good for LAL training.

From a different point of view, Gebril (2017) compared in-service and pre-service EFL teachers' conceptions about language assessment. He explored to what extent the assessment literacy activities given in training programs were effective. To address that issue, the researcher administered open-ended questions to 100 in-service teachers and 70 pre-service teachers who were EFL teachers or candidates in Egypt which relies on high-stakes exams. The findings indicated that while in-service teachers adopted both summative and formative assessment, pre-service teachers inclined to favor summative evaluation. In addition, in-service EFL teachers believed in the assessment function of reinforcing and improving learning whereas pre-service teachers were only interested in the assessment results. However, both groups of teachers held negative attitudes towards training programs due to their ineffectiveness.

Here is the summary of the research studies conducted in in-service EFL contexts regarding LAL within foreign contexts (see Table 2.1).

Table 2.1. *Research studies conducted in in-service EFL contexts about LAL within foreign contexts*

	Author(s)/Year	Context	Topic	Methodology	Results/Findings
1	Duboc (2009)	Brazil – EFL – in-service	conceptions & practices about LAL	- 3 elementary school EFL teachers (6 th grades) - interviews, observations, written tests	- evaluation = measurement - used linguistic code assessment (grammar & vocabulary) - employed written tests

2	Fulcher (2012)	worldwide – EFL – in- service	survey development about LAL in relation to assessment training needs	- 278 teachers from different countries - survey	- 4 factor-survey: test design, standardized testing, classroom testing, validity & reliability - expanded definition of LAL proposed - materials insufficient for training
3	Munoz, Palacio & Escobar (2012)	Colombia – EFL – in-service	beliefs, perceptions & practices about LAL	- 62 EFL teachers at language center of private university - surveys, written reports, interviews	- a mismatch between their beliefs and practices - believed in the usefulness of formative, but did use summative assessment
4	Wach (2012)	Poland – EFL – in-service	beliefs & practices about LAL in relation to being native or non-native speaker of English	- 87 EFL teachers (school & university) - open-ended questionnaires	- university teachers more aware - all focused on grammar & vocabulary assessment - being native or non- native did not affect
5	Razavipour (2013)	Iran – EFL – in-service	stakeholders’ views & needs of LAL in TEFL MA program	- 3 types of stakeholders: test developers, instructors, test- takers - survey - test samples by test developers collected	- test designers: knowledge & theory of assessment - instructors & students: skills of assessment - gap between test- designers and instructors & students
6	Tsagari (2013)	Cyprus – EFL – in-service	perceptions & practices about assessment in EFL classes	- 41 undergraduates from different subjects - questionnaire about the uses of assessment in their EFL classes	- most used: written tests, grammar & reading, multiple choice, essay - least used: portfolio, listening & speaking - summative than formative
7	Vogt & Tsagari (2014)	across Europe – EFL – in- service	assessment training needs of EFL teachers	- 853 EFL teachers from 7 European countries - questionnaire	- LAL did not develop - few received training & it did not prepare them much - no idea about recent assessment ways

8	Hakim (2015)	Saudi Arabia – EFL – in-service	awareness levels & practices about LAL in line with experience	- 30 female EFL instructors at university - questionnaire	- had adequate knowledge - but practices not developed - more experienced ones were more aware
9	Jannati (2015)	Iran – EFL – in-service	perceptions & practices about LAL in line with experience	- 18 EFL instructors at university - semi-structured interview	- all had good knowledge - did not reflect on practices - experience did not affect
10	Gebril, Boraei & Arrigoni (2016)	Egypt – EFL – in-service	training project for EFL educators about LAL	- 48 teachers attended the 4-day workshops about language assessment - reflective journals, plans of assessment, post-training surveys and follow-up activities collected	- positive development - training too intensive
11	Klinger (2016)	Colombia – EFL – in-service	conceptions & practices of assessment	- 4 EFL professors at university - phenomenology - interviews, documents	- most: to measure knowledge - appreciated formative assessment, but did not use
12	Y. Xu (2016)	China – EFL – in-service	planning assessment with respect to LAL	- 20 EFL teachers at university - lesson plans, interviews	- lesson plans included assessment - mismatch between learning goal and assessment task - intuitively planned assessments
13	Zolfaghari & Ahmadi (2016)	Iran – EFL – in-service	different academic backgrounds & LAL	- 32 high school teachers - 4 subjects: humanities, science, English language, and non-English language (Arabic/Persian) - semi-structured interview	- only two types: continuous/ diagnosis and continuous/final - familiar with, but used unconsciously - different academic backgrounds required different assessment
14	Berry, Sheehan & Munro (2017)	international in-service EFL teachers at British University	attitudes & needs about LAL	- 3 EFL teachers were observed and interviewed - 48 teachers for focus groups	- lacked training - needed materials to assess - not familiar with LAL concept

15	Baker & Riches (2017)	Haiti – EFL – in-service	in-service training to develop LAL	- 120 secondary school EFL teachers & also specialists - one-week workshop training - feedback on exams, surveys, interviews collected	- positive improvement - challenges in decision-making between teachers and specialists - specialists more knowledgeable
16	Djoub (2017)	worldwide – EFL/ESL – in-service	the effect of LAL levels on the views & practices	- 45 EFL teachers from different countries responded to a web-based questionnaire - 10 interviewed	- training did not assure to be literate - needed more autonomy, training, collaboration, feedback, materials
17	Hudaya (2017)	Indonesia – EFL – in-service	readiness, practices & problems about LAL	- 43 EFL teachers (different grades) - open-ended questions	- most felt ready to assess - practicality, reliability, validity, authenticity and washback important principles - difficulties in giving feedback
18	H. Xu (2017)	China – EFL – in-service	development of LAL	- 3 year-longitudinal study - 26 class observations, fieldnotes about 4 teachers, 48 interviews with 2 teachers, 171 journal entries	- three phases: 1. practical skills for planned assessment 2. more aware of the connection between teaching objectives and assessment 3. perform spontaneous assessment
19	Gebril (2017)	Egypt – EFL – in-service	comparison of pre- and in-service EFL teachers' conceptions about assessment	- 100 in-service & 70 pre-service teachers - open-ended questions	- in-service teachers adopted summative and formative assessment - pre-service teachers favored summative one - in-service teachers believed in the improvement side of assessment while pre-service teachers thought only assessment results - both held negative attitudes towards training due to their ineffectiveness

20	Sheehan & Munro (2017)	international in-service EFL teachers at British University	needs and materials about LAL training	- 3 EFL teachers interviewed and observed - 48 teachers for focus groups - 20 teachers' opinions in a workshop	- lack of confidence in testing - needed training in terms of practice - wanted ready-made materials to assess
21	Tsagari & Vogt (2017)	across Europe – EFL – in-service	perceptions & training needs about LAL	- 63 EFL teachers from 3 European countries - semi-structured interview	- most used: pen-and-paper tests, assessed grammar and vocabulary, marking only errors and mistakes - did not use alternative assessment - received little training - could not determine local needs
22	Y. Xu & Brown (2017)	China – EFL – in-service	demographic variables & LAL	- 891 EFL teachers - questionnaire	- level: basic - little effect of demographic variables on LAL
23	Fard & Tabatabaei (2018)	Iran – EFL – in-service	levels of LAL	- 52 EFL teachers - inventory	- low levels - not feel ready to assess
24	Giraldo Aristizabal (2018)	Colombia – EFL – in-service	beliefs & practices about LAL	- 60 EFL instructors at university - regarding the design of an achievement test, they responded to a questionnaire and were interviewed	- four principles required: reliability, validity, authenticity and washback - believed, but not reflect on practices
25	DiBiase-Lubrano (2018)	Yale's Center for Language Study	change of LAL after workshops at university	- 16 lectors - online survey	- gained much more knowledge about technical issues and reliability & validity - needed rating scale performance about speaking and writing assessment
26	Yan, Zhang & Fan (2018)	China – EFL – in-service	development of LAL	- 3 middle school EFL teachers - semi-structured interview	- developed LAL as a result of intuitions and experiences - context and experience affected LAL - needed training in terms of practice

So far, what was happening in foreign EFL contexts in terms of in-service teachers' LAL have been discussed. However, it is important to illustrate what the status of LAL in Turkey from the perspective of in-service EFL teachers, as well. As in foreign in-service contexts, there have been studies about the effect of demographic variables, beliefs, practices, needs and development of LAL in Turkey.

To begin with, Koksall (2004) examined what and how much English language teachers knew about testing and assessment. With this aim, he collected specific test samples of testing language skills from the state schools in which his ELT trainees were doing their practicum. A total of 56 different classroom test samples including quizzes, progress tests, and achievement tests were gathered from in-service EFL teachers. The analysis of the samples indicated that there were certain problems with testing. For example, the aim of assessment was not targeted to one skill or ability, some questions had more than one possible answer due to its not being contextualized, and some instructions were unclear and wrong. Moreover, some tests were like a collage of different ready-made tests, and most of tests were based on only recognition, not production. All the findings were connected to the insufficient training in testing because the researcher argued that testing and teaching did not overlap; they teach something but test something else.

Similarly, Saricoban (2011) analyzed the tests samples of in-service EFL teachers. He designed his study in order to scrutinize how English language testing was performed in state school from three points of views: Test construction, administration, evaluation and comparison of previous testing methods. For this aim, he collected test samples from in-service EFL teachers by the help of his teacher trainees who were doing their teaching practice at those schools. The analysis of test samples showed that most of the samples lacked content validity though they seemed to have face validity and reliability. The teachers tested only grammar and vocabulary, and sometimes reading and writing in a contextualized manner; there was no testing for listening and speaking skills. Therefore, tests generally demotivated students. The features of timing and scoring were not stated; however, the instructions were clear. Mostly, teachers used tests for the purpose of diagnosis. In sum, more training was needed to develop teachers' understanding and skills of assessment in different domains.

Differently, Oz (2014) investigated what Turkish EFL teachers preferred as assessment tools in their classes, and also, examined whether there was a difference in

assessment practices with regard some demographic variables. 120 EFL teachers working at state and private schools responded to an online Likert-type questionnaire about their practices. The results showed that many participant teachers preferred using traditional methods of assessment as summative evaluation rather than formative one, and they mostly utilized the techniques of multiple-choice tests, fill-in-the-blank items, true-false statements, matching items, and short-answer questions. Nevertheless, they rarely employed rubric, self/peer-assessment, observation, drama and the like. With respect to the demographic variables, only teaching experience significantly differed in monitoring activities. Therefore, he concluded that assessment training in teacher education programs may be revised to be helpful to change and affect their assessment knowledge and practices in the direction of more formative methods of assessment.

In addition, Buyukkarci (2016) examined the levels of LAL, and whether teaching experience and post-graduation had an impact on LAL of in-service EFL teachers. Conducting an assessment literacy inventory with 32 EFL teachers who were working at state schools and universities, the researcher found out that the sample had a low level of assessment literacy. Besides, neither the years of teaching nor post-graduation affected assessment literacy of EFL teachers. Hence, the researcher suggested that further training was required to enhance assessment literacy of EFL teachers.

In his another study, Buyukkarci (2014) explored the beliefs of primary school EFL teachers about formative assessment, and also, their formative assessment practices. 69 primary school EFL teachers rated a Likert-type scale about the beliefs and practices of formative assessment. Besides, 10 of them were interviewed to further analyze their uses and ideas of formative assessment in their classes. The findings showed that teachers had positive attitudes towards formative assessment. However, they did not employ formative assessment in their classes. They mostly relied on summative assessment tools owing to the crowded classes and their heavy work load.

Moreover, Sahinkarakas (2012) conducted a study to examine how teachers perceive assessment, and whether teaching experience has a role in the construction of language assessment beliefs. 53 pre-service EFL teachers and 47 in-service EFL teachers (half of them had more than 5 years of teaching experience, and the rest had less than 5 years of teaching) took part in the study. As a data collection tool, the researcher utilized a metaphor to reveal the conceptions of teachers. The analysis of the collected metaphors indicated that assessment was seen mostly as a formative tool by all the participants

whereas only pre-service teachers and some less experienced teachers also pointed out the issue of assessment as a sign of self-efficacy. In terms of assessment as a summative tool, mostly less experienced in-service teachers depended on using summative assessment. Therefore, the researcher concluded that both pre-service and in-service EFL teachers thought assessment is integrated in teaching, and teaching experience did not have any effect in assessment beliefs.

Doing a qualitative research, Saka (2016) investigated what EFL teachers thought about testing procedures which were practiced in state schools. To perform the study, she administered structured open-ended questions to 20 EFL teachers working at secondary schools of Turkey. The analysis of the questions yielded that the most utilized and preferred testing tool was multiple-choice test. Also, most of the teachers used written tests, and assessed reading and writing skills. Since parents expected high marks from their children, students were easily demotivated when they got lower marks in English tests. However, as they took higher scores, they were more motivated to learn English. According to the teachers, the biggest problem was that Turkey has a high-stakes-examination context, which influence the assessment practices. Therefore, although the participant teachers believed that such tests did not measure the real language performance of students, they continued to use multiple-choice tests owing to the testing context of Turkey.

In the same vein, Tuzcu-Eken (2016) explored what kind of problems about testing exist at vocational schools of university. To do so, she conducted semi-structured interviews with 5 EFL teachers about their views of testing and what kind of testing practices they implement. Also, two groups of students ($n=15$) studying at vocational school of the university took part in the focus group discussions. The findings showed that most of the participant students had lack of motivation, and their only goal of learning English was to pass the school. In contrast, teachers reported that they tried many testing techniques despite the lack of some facilities. They complained about physical conditions which were not suitable for, i.e., listening activities. They were also not satisfied with crowded classes because though they wanted to test speaking and listening, they could not do due to the mentioned problems.

Furthermore, Onalan and Karagul (2018) studied with EFL instructors to investigate whether their English-related undergraduate programs and teaching experience have an impact on their assessment practices. 70 EFL teachers working at a

preparatory school of a state university rated a Likert-type questionnaire about the beliefs about assessment by giving their background information. There were 51 ELT graduates, 15 English literature graduates, 2 translation graduates, and 2 linguistics graduates, and more than half of them were novice EFL teachers with one-year experience. The results of the questionnaire produced that most of EFL teachers believed in the effectiveness of using formative assessments. Especially, they thought self-assessment was very important. However, their undergraduate program types did not influence their assessment beliefs and preferences. Their teaching experience also did not affect their assessment procedures.

Similarly, Oz and Atay (2017) investigated the perceptions and practices of EFL instructors about in-class language assessment. 12 EFL instructors with different years of teaching experience and working at a preparatory school of a university reflected on their views and class practices of assessment through interviews. The findings yielded that there was a mismatch what they knew and believed and what they practiced in the classroom. Most of the participants were aware of basic types of classroom assessment, but they did not use them in their classes. In terms of experience, it did not have an impact on neither perceptions nor practices of EFL teachers about language assessment. The researchers concluded that EFL instructors could not reflect their LAL on their practices; so, in-service training programs may be helpful in that issue.

Regarding the practices of assessment as well as the grading procedures, Karagul, Yuksel and Altay (2017) carried out a cross-sectional study with different groups of EFL teachers. There were primary ($n=8$), secondary ($n=12$), high school ($n=2$), and university ($n=3$) EFL teachers as the participants, and then, a total of 25 EFL teachers took part in the research. An adapted questionnaire was used to elicit their assessment methods and techniques used in their classes. Although most EFL teachers regardless of their school type believed that alternative assessment such as project work and portfolio was a good way of assessment, they did not use them much in their classes. It is because their classes were very crowded and the class hours for English language were not enough. Thus, they complained about not using alternative assessment as much as possible. All in all, there was gap between the beliefs and practices of EFL teachers about language assessment.

On the other hand, Yastibas and Takkac (2018a) focused on the decision-making processes of EFL teachers about their assessment preparation process. Conducting a qualitative study, they examined how EFL instructors develop their assessment tools in

line with their course objectives. Think-aloud protocols from 8 EFL instructors working at a private university were collected. The analysis of the data revealed that EFL instructors paid attention to their coursebook structure while designing their language skill test items. The participants also emphasized that they tried to use assessment to leave positive washback on students' learning. Besides, they took care of certain principles such as reliability and validity while preparing language skills test items. In sum, the researchers indicated that EFL instructors were very dependent on their coursebook rather than course objectives.

In their another study, however, Yastibas and Takkac (2018b) investigated how EFL instructors develop their LAL. In fact, this study is based on Yastibas' (2018) doctoral dissertation in which he examined LAL of EFL instructors from different point of views. With respect to the development of LAL, a small-scale qualitative study was conducted with 8 EFL instructors working at university. As data collection tools, focus group and semi-structured interviews were performed. The findings indicated that three main factors had an impact on LAL development: 'Previous assessment experience, assessment training, and self-improvement'. Previous assessment experience was about their use and preparation of assessment methods while they were students, which shaped their beliefs of assessment. Assessment training was made up of not only pre-service education but also pedagogical formation courses and in-service training. All of such trainings influenced EFL instructors' LAL development while contributing their assessment knowledge. However, the most influential one was pre-service training because it provided EFL instructors with basic knowledge and skills with regard to language assessment. The last factor was self-improvement which indicated that EFL instructors gained experience of assessment while started to work as teachers, when they began to prepare and administer and evaluate their own exams, and accordingly, changed their instruction for better language teaching. Out of these three factors, the most influential one was self-improvement because as the researchers highlighted that EFL instructors gained experience while putting their knowledge into practice.

In terms of the needs of LAL, Mede and Atay (2017) designed a study to find out the needs of EFL instructors with respect to their LAL. They employed a questionnaire and focus group interviews to elicit LAL needs. 350 EFL instructors who had more than 5 years of teaching experience at language preparatory school of a university responded to the questionnaire. Also, 34 EFL instructors were interviewed about their perceptions

and practices about language assessment. The findings revealed that assessment training (pre- or in-service training), assessment needs, attitudes towards assessment, and current practices of assessment at their schools had an impact on their LAL. EFL teachers needed training the most in the area of classroom assessment because they lacked some skills such as constructing language tests and giving feedback. They also had difficulty in testing language skills, but they were able to tests grammar and vocabulary properly. EFL instructors reported that pre-service training of language assessment was not sufficient, it only provided them with the knowledge of summative evaluation. Thus, the results showed that more assessment training was needed to develop assessment skills of EFL instructors, especially, in terms of classroom-assessment and testing language skills.

Likewise, Semiz and Odabas (2016) investigated the needs of EFL teachers about their language assessment; and also, the familiarity with assessment procedures were examined. An adapted questionnaire was utilized to find out LAL needs of in-service EFL teachers. There were 48 EFL teachers working at primary, secondary and high school who filled the questionnaire. The results yielded that most of the participants were familiar with assessment methods. But they reported that the training to understand what assessment literacy means was needed. Pertaining to the use of assessment, they mostly employed grammar and vocabulary tests purposefully, and they believed they were able to prepare language tests. However, they did not much use portfolios, statistics to analyze the results, and not consider the issues of reliability and validity. Therefore, the researchers found that more training about formative assessment, and about testing other language skills were needed.

Another study which is based on the investigation of LAL needs was performed by Turk (2018) with EFL instructors. He explored the perceived training needs with respect to some variables such as major, qualification degree in education and responsibility in testing offices. A total of 246 EFL instructors who were working at preparatory schools of Turkish universities participated in the study. Designing a mixed-methods study, the researcher used questionnaires for 246 instructors and interviews for 12 instructors. The analysis of the data showed that EFL instructors were familiar and seemed to be trained in most of the language testing and assessment areas. However, the sample indicated that they needed training in formative assessment as well as the concepts of language assessment. From the point of teaching experience, less experienced instructors tended to

need more training whereas more experienced reported they did not need professional development about LAL at all.

Here is the summary of the research studies conducted in in-service EFL contexts regarding LAL in Turkish context (see Table 2.2).

Table 2.2. *Research studies conducted in in-service EFL contexts about LAL in Turkish context*

	Author(s)/Year	Context	Topic	Methodology	Results/Findings
1	Koksal (2004)	Turkey – EFL – in-service	knowledge of language assessment	56 different test samples of teachers were collected	- problems: aim of the test not targeted to one skill; more than one possible answer; use of ready-made tests; only recognition tests
2	Saricoban (2011)	Turkey – EFL – in-service	practices of testing	test samples of teachers were collected	- problems: no content validity; only grammar and vocabulary testing; no testing listening and speaking; timing and scoring not stated; only diagnosis purpose - positive sides: face validity and reliability; clear instructions
3	Sahinkarakas (2012)	Turkey – EFL – in-service & pre-service	perceptions of assessment with regard to experience	- 53 pre-service and 47 in-service EFL teachers - metaphors collected	- assessment = formative - only pre-service and less experienced in-service teachers saw assessment as a sign of self-efficacy - less experienced teachers used summative more - experience did not much affect
4	Buyukkarci (2014)	Turkey – EFL – in-service	beliefs about formative assessment	- 69 primary school teachers - Likert-type scale - 10 of them were interviewed	- positive attitudes towards formative assessment - but did not use it - only used summative assessment
5	Oz (2014)	Turkey – EFL – in-service	preferences and uses of assessment tools in relation to assessment for learning and demographic variables	- 120 EFL teachers from state and private schools - Likert type questionnaire	- mostly, traditional methods and summative evaluation - most: multiple-choice, filling blanks, true-false, etc. - rarely: rubric, self/peer assessment, observation, etc. - teaching experience affected preferences and uses

6	Buyukkarci (2016)	Turkey – EFL – in-service	levels of LAL in line with experience and post-graduation	- 32 EFL teachers from state schools and universities - assessment literacy inventory	- low level of LAL - neither the years of teaching nor post-graduation affected LAL
7	Saka (2016)	Turkey – EFL – in-service	opinions about testing procedures in state schools	- 20 EFL teachers from secondary schools - open-ended questions	- most utilized & preferred: multiple-choice test - most: written exams - parents' high expectations demotivated Ss in English testing - problem: Turkey = high-stakes testing
8	Semiz & Odabas (2016)	Turkey – EFL – in-service	needs about LAL and familiarity with assessment	- 48 EFL teachers working at primary, secondary and high school - questionnaire	- most familiar - need to understand what LAL means - mostly employed grammar and vocabulary tests - did not use portfolios, reliability and validity, statistics
9	Tuzcu-Eken (2016)	Turkey – EFL – in-service	problems of testing at university vocational schools	- 5 EFL teachers - 15 Ss - semi-structured interviews	- Ss lack of motivation; only tried to pass English - Ts complained about physical conditions regarding testing; not satisfied with crowded classes
10	Karagul, Yuksel & Altay (2017)	Turkey – EFL – in-service	practices of assessment and grading procedures	- a total of 25 EFL teachers from different settings (primary, secondary, high school, university) - questionnaire	- all believed alternative assessment good, but did not use it - reasons: crowded classes and inadequate class hours - mismatch between beliefs and practices
11	Mede & Atay (2017)	Turkey – EFL – in-service	needs, perceptions and practices about LAL	- 350 EFL instructors responded to the questionnaire. - 34 EFL instructors were interviewed	- assessment, needs, attitudes and current practices of assessment affected LAL - needed training the most in the area of classroom assessment - able to test grammar and vocabulary, but difficulty in other skills - pre-service assessment training not sufficient

12	Oz & Atay (2017)	Turkey – EFL – in-service	perceptions and practices about in-class language assessment	- 12 EFL instructors from preparatory school - interview	- mismatch between knowledge/belief and practice - most aware of basic types of classroom assessment, but did not use them - experience = no effect in assessment practices
13	Onalan & Karagul (2018)	Turkey – EFL – in-service	the effect of undergraduate program and teaching experience in assessment practices	- 70 EFL instructors from preparatory school - Likert-type questionnaire	- 51 ELT; 15 English literature; 2 translation; 2 linguistics graduates - more than half of them novice (1 year of teaching only) - most believed in the usefulness of formative assessment - both undergraduate program and experience did not affect assessment practices
14	Turk (2018)	Turkey – EFL – in-service	needs about LAL	- 246 EFL instructors from preparatory school (questionnaire) - 12 EFL instructors for interviews	- familiar with and trained in assessment - needed training in formative assessment - less experienced instructors tended to need more training
15	Yastibas & Takkac (2018a)	Turkey – EFL – in-service	decision-making processes about assessment preparation process	- 8 EFL instructors working at a private university - think-aloud protocols	- paid attention to their coursebook structure; so, not dependent on course objectives - tried to use assessment to leave positive washback - cared reliability and validity issues
16	Yastibas & Takkac (2018b)	Turkey – EFL – in-service	development of LAL	- 8 EFL instructors working at a private university - focus group and semi-structured interviews	- three factors: ‘Previous assessment experience, assessment training, and self-improvement’ - the most influential one was pre-service training

In conclusion, all the studies conducted in in-service EFL contexts contributed to the LAL literature in terms of different significant points such as beliefs, attitudes, practices, needs and development of LAL.

2.3.4.2. Research studies conducted in pre-service EFL contexts

The research into the LAL of pre-service EFL teachers has just begun to flourish. It is because there are more studies conducted with in-service EFL teachers compared to pre-service teachers in foreign language education. The research studies about pre-service EFL contexts have focused on the experiences, beliefs, practices, and mostly, the needs of teacher candidates.

Regarding the experiences of formative assessment, Howerton (2016) designed a study to explore what kind of formative assessment was undertaken in teaching practice. Based on an exploratory qualitative study, the researcher collected information about the experiences of formative assessment by means of questionnaire, observation and interview. 5 pre-service teachers as high school English teacher candidates took part in the study. The study was executed during their teaching practice in which both their instructional practices and formative assessment were examined. The analysis of data produced three key findings. Firstly, teacher candidates began to put their theoretical knowledge into practice by doing assessments. Secondly, in this way, they learned assessment best by combining both knowledge and skills, which indicates assessment should be practiced even in pre-service years. Lastly, training needs about formative assessment were about knowledge, purpose and content standards; so, more emphasis on knowledge base, content area information, and more purposeful assessment should be highlighted in pre-service training with respect to formative assessment.

Similarly, Prasetyo (2018) investigated the development of LAL during teaching practice. As a one-case study, the researcher carried out the study with 1 student-teacher of ELT who finished his/her teaching practice in Indonesia. The teacher candidate was interviewed about his/her assessment experiences in the practicum. The findings indicated that the participant did not receive any support or help from his/her supervisor and cooperating teachers about language assessment. Besides, s/he only administered the assessment tasks given by his/her headmaster in the school. Therefore, s/he developed a sense that assessment was just about administering tests. In addition, s/he did not know how to prepare a language test or a rubric to assess. However, the participant only did sort of formative assessment by doing quizzes, asking comprehension questions and giving assignments to students during the practicum. Thus, the researcher highlighted there was a need to establish a platform in which EFL teacher candidates would gain more experiential learning in terms of assessment in their teaching practice.

Addressing the assessment training needs of pre-service teachers, Volante and Fazio (2007) investigated the needs of assessment literacy of teacher candidates. Though they studied with pre-service teachers in general, their study has significant contributions to the EFL context. As the participants, there were 69 pre-service teachers majoring in primary/junior education in Canada. They responded to a questionnaire about the level of assessment literacy, purposes and uses of assessment, and needs of training. The results showed that their assessment literacy levels were too low, and most of them used assessment for the purposes of summative evaluation. Also, most participants indicated that they needed training in the area of authentic assessment, specifically, portfolio and performance assessment. Lastly, they suggested that a specific course designed within the scope of classroom assessment needed to develop assessment literacy.

Likewise, Viengsang (2016) was concerned with LAL of pre-service EFL teachers. She carried out her study in order to investigate what teachers understood regarding assessments, what kind of experiences about language assessment they had previously, and what they did in terms of assessment in their practicum. A sample of 46 Thai pre-service EFL teachers who were doing their teaching practice at public and private schools took part in the study. The sample responded to a questionnaire about LAL, and then, 5 of them were interviewed for further understanding of the concept. The findings revealed that the assessment training provided them with a basis for language assessment; nonetheless, the teacher candidates could not perform assessment in the practicum. They also had some misunderstandings about test choices, approaches and principles. Hence, the training could not supply enough practical assessment for teacher candidates.

Here is the summary of the research studies conducted in pre-service EFL contexts regarding LAL within foreign contexts (see Table 2.3).

Table 2.3. *Research studies conducted in pre-service EFL contexts about LAL in foreign contexts*

	Author(s)/Year	Context	Topic	Methodology	Results/Findings
1	Volante & Fazio (2007)	Canada – pre-service general	assessment training needs about LAL	- 69 pre-service teachers majoring in primary/junior education - questionnaire	- assessment literacy levels were too low - most used summative one - most needed training in authentic assessment (portfolio and performance assessment)

2	Howerton (2016)	the USA – ESL – pre-service	formative assessment during teaching experience	- 5 pre-service teachers as high school English teacher candidates - questionnaire, observation and interview	- began to put their theoretical knowledge into practice - needs about formative assessment training: knowledge, purpose and content standards
3	Viengsang (2016)	Thailand – EFL – pre-service	conceptions and experiences about assessment in practicum	- 46 pre-service EFL teachers - questionnaire - 5 of them interviewed	- training provided them with basis of assessment - could not perform assessment in the practicum - some misunderstandings about test choices, approaches and principles
4	Prasetyo (2018)	Indonesia – EFL – pre-service	development of LAL during teaching practice	- one pre-service teacher of ELT (one-case study) - interview	- did not receive any support or help about language assessment - assessment = administrating tests only - did not know how to design and score tests - only did sort of formative assessment by doing quizzes, asking comprehension questions, etc.

As far as Turkish EFL context is taken into account, there have been recent but few attempts to study on LAL, especially, with regard to beliefs and preferences within the scope of pre-service foreign language teacher education.

Karaman and Sahin (2014) explored what teacher candidates thought about assessment and what the levels of assessment literacy they had. To do so, they utilized three instruments: An assessment literacy inventory consisting of assessment scenarios, an assessment literacy thoughts scale, and an assessment literacy attitudes scale. 289 pre-service teachers majoring in classroom, science, Turkish, English, computer, history, and geography teaching subjects responded to the scales. The data were analyzed via descriptive statistics. The results yielded that most participants needed more help to enhance their knowledge and skills about assessment since their levels in assessment literacy were low. Also, most participants thought that assessment was used to improve learning, and supported constructivist paradigms in assessment.

Moreover, C. Yetkin (2015) investigated the perceptions of EFL teacher candidates about how language assessment was used in primary settings. By using an adapted

questionnaire about assessment literacy, the researcher surveyed pre-service teachers regarding the topics of purposes, uses, approaches, and needs of classroom assessment. 30 pre-service EFL teachers filled the questionnaires. The results indicated that the participants used assessment with the purpose of observing the development of students. Thus, the most favored assessment technique was selected as observation technique the most. Most of the participants also favored performance and portfolio assessment within classroom assessment. Though the sample found the course was helpful, they indicated that they needed training more in other assessment strategies.

Furthermore, R. Yetkin (2017) was concerned of the conceptions of pre-service ELT teachers about language assessment in relation to some variables such as gender, age, and years of learning English. In order to explore their conceptions of assessment, a kind of assessment inventory was administered to 204 pre-service English teachers who were sophomore, junior and senior undergraduates in the ELT program. The results showed that the most rated purpose of assessment was about the improvement of learning, and assessment was believed to give feedback about students' performance. Among the demographic variables, only grade level (being sophomore, junior or senior) made a significant difference regarding the conceptions of assessment.

In addition, Bayram (2015) examined the preferences of pre-service EFL teachers about language assessment. The researcher also tried to elicit why they preferred certain types of assessment compared to others. Therefore, 326 pre-service EFL teachers completed assessment awareness, techniques usage and preferences questionnaires. Besides, 38 teacher candidates were interviewed about their preferences and reasons to choose some assessment types than others. The results highlighted that pre-service teachers were aware of different assessment techniques; and hence, they differed in terms of their preferences according to gender and school year. However, most of them favored to use translation and presentation to assess in the future while alternative assessment methods were not mentioned. The factors affecting their preferences of assessment were about the dominance of high-stakes tests, crowded classes, test anxiety and prior knowledge.

Finally, Komur (2018) conducted a study to find out how much pre-service EFL teachers were aware of assessment, and whether they were ready to perform assessment practices. Doing a qualitative case study, the researcher utilized open-ended questions to collect the data. 49 last year undergraduate students who would be soon graduated from

the program answered the questions. The findings revealed that though most of the participants were aware of skills and competencies of ELT, they underlined that they needed more training in terms of practicing assessment tools. They also regarded speaking skill was very important to be assessed. They seemed to feel ready for assessment due to the acquired knowledge and skills, but they pointed out that they were not ready for actual classroom assessment owing to lack of experience in the field. Hence, more practice-based LAL training was needed.

Here is the summary of the research studies conducted in pre-service EFL contexts regarding LAL in Turkish context (see Table 2.4).

Table 2.4. *Research studies conducted in pre-service EFL contexts about LAL in Turkish context*

	Author(s)/Year	Context	Topic	Methodology	Results/Findings
1	Karaman & Sahin (2014)	Turkey – pre-service general	opinions about assessment and levels of LAL	- 289 pre-service teachers majoring in classroom, science, Turkish, <i>English</i> , computer, history, and geography teaching subjects - assessment literacy inventory	- levels of LAL too low - needed training in knowledge and skills of assessment - most believed in assessment for learning
2	Bayram (2015)	Turkey – EFL – pre-service	preferences of assessment	- 326 pre-service EFL teachers for questionnaire - 38 pre-service teachers for interviews	- aware of different assessment techniques - differed in terms of their preferences according to gender and school year - them favored to use translation and presentation to assess - the factors: dominance of high-stakes tests, crowded classes, test anxiety and prior knowledge
3	C. Yetkin (2015)	Turkey – EFL – pre-service	perceptions & practices of assessment in primary school	- 30 pre-service teachers - questionnaire	- used assessment to observe Ss - favored performance and portfolio assessment most - needed training more in other assessment strategies
4	R. Yetkin (2017)	Turkey – EFL – pre-service	conceptions of assessment in relation to demographic variables	- 204 pre-service English teachers (sophomore, junior and senior) - assessment inventory	- most: purpose of assessment = to improve learning & give feedback - only grade level (being sophomore, junior or senior) made a significant difference regarding the conceptions

5	Komur (2018)	Turkey – EFL – pre- service	awareness and readiness about assessment	- 49 last year undergraduate students - open-ended questions	- aware of skills and competencies of ELT - needed more training in terms of practicing assessment tools - seemed to feel ready, but not ready for actual classroom assessment
---	--------------	--------------------------------------	---	--	---

As a conclusion, there is a need to focus on pre-service EFL teachers in terms of LAL because though there have been significant contributions about needs, beliefs, and experiences in the practicum, there have still been gaps for improvement, especially about the training for a better future teaching life.

2.4. Training of Language Assessment Literacy (LAL)

2.4.1. Importance of LAL training

Language assessment is an indispensable component of foreign language education. It is because assessment is helpful to show how much progress is made in language learning, to organize teaching practices for better learning opportunities and motivating students to learn the target language, and to collect data about both teachers' and students' developments (Oz & Atay, 2017). Language assessment is also important to reduce the negative feelings and attitudes towards testing while creating enthusiasm and motivation for learning by means of positive impact (Watanabe, 2011). Then, to perform sound assessment so as to promote language learning, teachers must possess LAL to carry out the assessment process effectively. To help language teachers to be assessment literate, there are some suggested ways to be fulfilled. For example, language teachers need to develop an understanding of what sound and unsound assessment are, to gain knowledge and skills of performing assessment methods, and to recognize and be aware of different approaches and techniques available for use to do assessment according to their own instructional contexts (Djoub, 2017).

In order to provide language teachers to develop their LAL, either pre-service courses about language assessment given at teacher education or in-service training workshops or seminars given as a part of professional development may be helpful. However, as the literature has informed that there are in-service EFL teachers who are not able to assess their students effectively because they are not assessment literate due to not taking assessment courses at the level of teacher training. Then, it makes LAL

training a must in this sense, for teachers are acknowledged to be able to teach and assess at the same time. Thus, one of the ways to promote LAL of foreign language teachers is through assessment courses given in training years (Jeong, 2013). It is important to give LAL training early in learning owing to the fact that language teachers' knowledge, skills, beliefs and abilities are shaped in teacher education mostly, and it might be too late to give LAL training via in-service training. It is also important to give LAL training in teacher education programs before starting to work professionally as language teachers because assessment is embedded in teaching, and it cannot be considered as a separate feature of teacher career. Although there is no guarantee in terms of gaining necessary knowledge and skills about language assessment via pre-service courses (Hudaya, 2007), it is useful to introduce pre-service teachers with language assessment in teacher education and make them ready for their future assessment practices. However, it should be noted that pre-service teachers might develop their LAL more with the help of in-service training by acquiring more knowledge and experience about assessment while carrying on their teaching.

Considering the LAL training in teacher education programs, a fixed implementation of course content across different contexts cannot be expected because each context has its own characteristics. However, most language assessment courses have been given in a one-size-fits-all-manner by focusing on mostly summative assessment and testing perspective (Gebriel, Boraei & Arrigoni, 2016). Therefore, there is a need to design and develop LAL courses in teacher training to respond to the contextual features and its course-takers' needs for future appropriate practices (Berger, 2012; Jeong, 2013). Even though there cannot be a standard about how language assessment courses in pre-service teacher education are taught, they must have certain aspects to equip their language teachers with required knowledge and skills of LAL. Some researchers have discussed that language assessment courses should have a balanced content covering what assessment means, how assessment is carried out, why it is important to assess, how each language skill and area can be tested and evaluated, and how the results are interpreted to give feedback to enhance and motivate learning (Berger, 2012; Huang & He, 2016; Inbar-Lourie, 2008). Besides, language teachers must be aware of the fact assessment is integrated into teaching for better learning and teaching by the help of such training (O'Loughlin, 2006). Moreover, through language assessment courses, the required LAL components can be supplied such as identifying and using suitable assessments in line

with the purposes and contextual factors, analyzing assessment results to improve teaching as well as learning, and developing an understanding of sound and integrated assessment for pedagogical purposes (Berry & O’Sullivan, 2016). Drawing attention to the fact that language assessment courses are peculiar to their own contexts, Taylor (2009) has argued that;

‘training for LAL entails an appropriate balance of technical know-how, practical skills, theoretical knowledge, and understanding of principles, but all firmly contextualized within a sound understanding of the role and function of assessment within education and society. The exact nature of this balance and the level of literacy to be acquired may vary according to context, and according to whether test stakeholder constituencies are more or less directly involved in matters of test selection, test production, test score interpretation, or test evaluation’ (p. 27).

As a conclusion, it can be maintained that LAL training in teacher education programs are very important to enable future teachers to understand and to be able to practice effective assessment to improve learning as well as to motivate language learning in their professional lives.

2.4.2. LAL training in EFL/ESL contexts across the world

The perspective of language assessment has been changed from a testing notion to an assessment viewpoint recently due to the changes in the perceptions and implementations of foreign language teaching methodologies. Therefore, language assessment has not been recognized as a technical and psychometric knowledge area anymore; however, it is acknowledged to be a social practice including political and ethical issues due to the fact that assessment forms a basis for improving learning and teaching at the contextual level (O’Loughlin, 2006). Hence, language assessment courses have been constructed in line with such considerations including more classroom-based assessment topics, which underlines that teaching and testing are integrated (Montee, Bach, Donovan & Thompson, 2013). However, how language courses have been practiced and what should be done to improve their content are the recent concerns of most researchers within the scope of LAL training. Since the present study is based on pre-service teacher education context, despite their significant contributions, the studies about language assessment courses given as postgraduate courses, MA courses, or general education assessment courses (i.e., DeLuca & Klinger; 2008; Kleinsasser, 2015; O’Loughlin, 2006; Orr, 2010) are excluded from the discussion.

Brown and Bailey (2008) investigated the characteristics of language testing courses in teacher training, and whether there were changes over 10 years of teaching such courses (from 1996 to 2007). Thus, they examined how those courses were carried out in terms of instructors, content, and students by means of a questionnaire. 97 participants from different nations responded to the online questionnaire. The results showed that most of the students held positive beliefs about language testing after they took the course though they had negative feelings towards testing before taking the course. Also, course instructors improved themselves more in the topics of test critiquing and test analysis over the years. Thus, they included such topics and focused more on item writing, score interpretation, critiquing and analysis of tests in their language testing courses. Although there were changes in language testing courses, most of the content remained relatively stable over the years.

Moreover, Jin (2010) designed a study to describe how language testing and assessment courses were executed in ELT training programs in China. To elicit the content of language testing courses, the researcher constructed a questionnaire to investigate the perceptions, teaching content, method, and materials used in those courses. Language testing courses offered in five disciplines 'Curriculum and Teaching Methodology, Master of Education, Linguistics and Applied Linguistics, English Language and Literature, Foreign Linguistics and Applied Linguistics' were evaluated in that sense. 86 course instructors responded to the questionnaire, and the results revealed that language testing courses included basic theoretical issues about assessment whereas educational and psychological measurement as well as classroom-assessment were emphasized less in those courses. Besides, the practical side of classroom assessment and statistical analyses for testing were not covered in the courses. The participant instructors also stated that their students were motivated in those courses, but they expected more practice than theory. The commonly used testing books belonged to Bachman, Hughes, Heaton, and Bachman and Palmer. Lastly, the name of those courses inclined to reflect testing perspective than assessment.

Likewise, Csépes (2013) focused on language testing and assessment course given in ELT programs as a degree of MA in Hungary. The course content was comprised of basic concepts of language testing, types of tests, testing language skills and areas. The teaching mode in that course included hands-on-experience, simulations, tutor's guide and feedback. The effect of the course was explored through the perceptions of pre-

service and in-service teachers. The questionnaires were distributed before and after taking the course. 15 pre-service teachers and 17 in-service teachers filled the questionnaires. The overall finding produced positive views and changing attitudes about language testing course. Regardless of teacher types as pre-service and in-service, most of them liked the topics of alternative assessment, assessing speaking and writing as well as grammar and vocabulary. Also, most of them stated that their perceptions changed, especially in the areas of analyzing test data, using self-assessment and portfolio. They were able to design and choose suitable assessment methods after the course. Besides, sociocultural context, course lecturers' beliefs and experiences, and educational policy had a great impact on their LAL construction. Hence, the course was proved to be useful both for pre-service and in-service EFL teachers.

In the same vein, Lam (2014) designed a comprehensive research which aimed to investigate the overall language assessment training programs offered in five Hong Kong English language teacher education institutions. More specifically, whether such programs were facilitative or debilitative in the development of teachers' language assessment literacy. A survey of the programs and government documents, interviews with selected pre-service teachers and instructors, the evaluation of students' assessment tasks, and the evaluation of assessment courses were utilized to investigate the effect of the programs. Five main themes were reported: The assessment courses in the programs were not effective in enhancing LAL; the application of LAL in real classroom contexts was limited; there was a lack of training in administering assessments of both classroom-based and large-scale; the experiences of the participants in the course were varied, either being positive or negative; the perceptions related assessment were examination-oriented due to their exam-oriented culture. Therefore, in general, the assessment courses were not as effective as expected in developing LAL of its pre-service language teachers.

Watanabe (2011) explored how a test literacy course in general education given at university was evaluated by its course-takers in Japan. The objectives of that course were to provide undergraduates with basic principles of language testing, changes in the attitudes towards testing, and the ability to make and use principled language tests. 113 undergraduate students majoring in education and taking EFL classes took part in the study. They wrote essays about the course of test literacy before and after the course. The analysis of their essays before taking the course yielded that they expected ideal testing practice, and mentioned their previous negative experiences with language tests,

especially, high-stakes tests. Therefore, they had negative attitudes towards testing before the course. The other data about the impressions of students after the course produced that their views about testing changed in a more positive way, they became more knowledgeable about testing principles and criteria, and however, they had challenges in marking or scoring language tests.

In addition, Jeong (2013) scrutinized the effect of course instructors on language assessment courses in teacher training programs from the point of language testers and non-language testers. Language testers were the experts in language testing whereas non-language testers were the experts in language teaching methodologies as well as having some experiences with assessment activities. Conducting online surveys and interviews, the researcher compared those two samples in terms of their language assessment courses. 140 instructors responded to the survey, and 13 instructors were interviewed. The survey results showed that there were not many differences between the content of language assessment courses between language testers and non-language testers. While language-testers were more confident in doing and using tests technically, non-language testers tended to use more classroom-based assessment rather than technical testing. Therefore, language-testers spent more time on the topics of tests theory whereas non-language testers concentrated on classroom assessment topics the most in their courses. However, the course structures were similar for both groups of participants. In short, language testing courses were more or less similar both for language-testers and non-language testers despite different background of the participants. The only difference was about their content focus; test theory vs classroom assessment topics.

Giraldo and Murcia (2018) conducted an action research to find out what should be covered in a language assessment course from the eyes of its stakeholders in a pre-service EFL program. The action research was undertaken in the program of bilingualism with a focus on ELT in Colombia. Language assessment course titled ‘Seminar in Language Assessment’ was added to the last semester of the program and delivered 4 hours a week. In order to diagnose the assessment training needs to shape the course, 5 professors completed a questionnaire, 1 expert of language education was interviewed, and 30 pre-service teachers responded to a questionnaire in which their expectations about the course were interrogated. The findings indicated that pre-service teachers expected to learn how language skills were tested and assessed whereas they were not interested in test specifications, administration and ethical concerns. All the participants as the

stakeholders perceived language assessment as the combination of theory and practice. They indicated that the course should cover the relationship between language assessment and teaching methodologies, as well. Finally, the sample expected to gain information about Colombian general policies about assessment.

Similarly, Ukrayinska (2018) designed a study to describe the framework for LAL training in Ukraine, and to explore how the course can be improved more in terms of its content. The name of the course was 'Introduction to Testing' taught at a university to promote LAL of pre-service EFL teachers. Nearly 150 student teachers who were taking the course participated in the study. Questionnaires before and after the course were distributed, and evaluations of assessment tasks were collected. The analysis of the gathered data showed that student teachers had many challenges in designing language assessments; for example, they could not construct their tests according to the appropriate proficiency level described in CEFR. Although they recognized which language skill was tested by evaluating the assessment tasks, they were not able to identify which subskills were tested. Therefore, the course content should be developed more to respond to the needs of the student teachers.

Focusing on student teachers' beliefs and conceptions of LAL, Hilden and Frojdendahl (2018) executed their study to find out what Finnish pre-service EFL teachers thought before taking the course and after the course about language assessment. Before the language assessment course, 77 student-teachers responded to a web-based survey. After the course ended, 65 teacher candidates filled the survey. The survey was about the attitudes and practices about the domains of language assessment. The results revealed that the participants gained certain abilities related to language assessment throughout the course. They also developed their conceptions of assessment towards more learner-centered forms. Besides, the participants changed their attitudes and planned assessment into more working skills. All in all, the language assessment training course was proved to be beneficial to teacher candidates in terms of transforming them into more knowledgeable and skillful teachers in assessment.

Different from EFL assessment courses, within the context of teacher training programs, Montee, Bach, Donovan, and Thompson (2013) examined the assessment needs of 29 pre-service teachers of 'Less Commonly Taught Languages (LCTLs)'. The sample was made up of the teachers of 'Arabic, Chinese, Dari, Hindi, Persian, Swahili, Turkish, and Urdu'. The participants enrolled in a seven-week course in language

assessment, and their needs were investigated with the help of a questionnaire before the course and the assessment tasks developed by participants during the course. Their perceptions of course instructors were also inquired. The findings showed that before the course, the sample was not familiar with the concepts and terminology of assessment. However, at the end of the course, they were able to link assessment and instruction in a meaningful way, and both the knowledge and the awareness about assessment were increased. Therefore, the researchers concluded that such a training course was effective in improving assessment literacy of LCTLs.

Here is the summary of the research studies conducted in EFL/ESL contexts across the world regarding LAL training (see Table 2.5).

Table 2.5. Research studies conducted in EFL/ESL contexts across the world about LAL training

	Author(s)/Year	Context	Topic	Methodology	Results/Findings
1	Bailey & Brown (2008)	worldwide – EFL/ESL – training	characteristics of language testing courses in teacher training (from 1996 to 2007)	- 97 participants from different nations - online questionnaire	- positive beliefs after course - course instructors improved in test critique and analysis more - topics added: item writing, score interpretation, analysis of tests - most content remained stable over the years
2	Jin (2010)	China – EFL – training	how ELTE course was executed	- 86 course instructors - questionnaire	- included basic theoretical issues - less classroom assessment and psychological measurement - practical issues and statistical analyses not included - Ss motivated, expected more practice - commonly used book: Heaton, Hughes, Bachman, etc.
3	Watanabe (2011)	Japan – EFL – training	a test literacy course in general education in relation to EFL classes	- 113 undergraduate students majoring in education and taking EFL classes - essays before and after the course	- before: negative attitudes, expected ideal testing practice, previous negative testing experiences - after: positive changes in the views, more knowledgeable, challenges in marking and scoring

4	Csépes (2013)	Hungary – EFL – training	language testing and assessment course given in ELT programs as a degree of MA	- 15 pre-service & 17 in-service teachers - questionnaire	- positive views and changing attitudes = course beneficial - changed perceptions regarding analyzing tests, using self-assessment and portfolio - able to design and choose suitable assessment methods - sociocultural context, course lecturers' beliefs and experiences, and educational policy had a great impact on their LAL
5	Jeong (2013)	South Korea – EFL – training	the effect of course instructors on language assessment courses: language testers vs non- language testers	- language tester (LT)= expert - non-language tester (non-LT) = regular teacher - 140 instructors for questionnaire - 13 instructors for interview	- not many differences between two groups in terms of content and course structure - LTs = doing and using tests technically; non-LTs = used more classroom-based assessment - only content focus differed
6	Montee, Bach, Donovan & Thompson (2013)	other languages' assessment training	needs of assessment training in the program of 'Less Commonly Taught Languages'	- 29 pre-service teachers majoring in teaching different languages - questionnaire :before & after - assessment tasks collected	- before: not familiar with assessment - after: linked assessment and instruction; effective training in assessment literacy
7	Lam (2014)	China – EFL – training	overall language assessment training programs	- 5 Hong Kong English language teacher education institutions - documents, interviews with pre-service and course lecturers	- courses not effective in LAL development - practice limited - lack of training about large-scale and classroom testing - positive or negative experiences of participants - exam-oriented perceptions dominant

8	Giraldo & Murcia (2018)	Colombia – EFL – training	language assessment course evaluation in the program of bilingualism	- 5 professors for questionnaire - 1 expert interviewed - 30 pre-service teachers for questionnaire	- pre-service teachers expected to learn how language skills were tested - all: assessment = theory + practice - course should cover the relationship between language assessment and teaching methodologies - expected to gain info about Colombian general policies about assessment
9	Hilden & Frojdendahl (2018)	Finland – EFL – training	beliefs and conceptions of LAL	- 77 student teachers: before - 65 student teachers: after - web-based survey: before & after	- gained certain abilities related to language assessment - developed more learner-centered assessment conceptions - attitudes changed; course beneficial
10	Ukrayinska (2018)	Ukraine – EFL – training	describe the framework for LAL training	- nearly 150 student teachers taking the course - questionnaire before & after - evaluation tasks collected	- challenges in designing, especially, according to level - tested language skill detected but could not identify subskills to be tested

In conclusion, most recently, there have been significant attempts to describe LAL training in pre-service teacher education across the world within their own unique contexts. Though there are similar or different findings and practices, all the language assessment courses aim to improve LAL of ELT teacher candidates.

2.4.3. LAL training in Turkey

In Turkey, ELT teacher education has been given by Faculty of Education at universities as an undergraduate program for teacher education. In this ELT training program, undergraduate students take English related courses such as language skills (reading, writing, listening, speaking) and areas (grammar, vocabulary), English literature and linguistics, translation, material development, methods courses like teaching grammar, speaking, listening, reading and writing, and foreign language assessment, all of which are given in English language. In addition to such courses, students take pedagogical courses in terms of general education given in Turkish language as well as some elective courses to graduate from the program. They also must do teaching practice

at their last year of the program lasting two semesters before starting to work as EFL teachers across different school levels ranging from primary to high school, and even at universities as instructors. The entire ELT teacher education is made up of 4-year training.

Considering the language assessment course in pre-service ELT training program, it has been offered since 1998 (Koksal, 2004). It is an important course for Turkey because high-stakes testing is dominantly used especially as entrance exams in the country. Thus, exams and assessment are very crucial for students to place in schools as well as to improve themselves as better learners, which makes pre-service teachers equip with necessary knowledge and skills of language assessment for their professional teaching lives. Although there is another course called ‘Measurement and Assessment in Education’ given in the sixth semester in Turkish as a general education course, ‘English Language Testing and Evaluation (ELTE)’ course is a domain-oriented course which is specifically about how to assess language skills and areas in English. It is delivered 3 hours a week in the last semester of the program in English language as a compulsory course before pre-service teachers are graduated from the program and after they take all the teaching methods courses.

2.4.3.1. Content of LAL training in Turkey

The course of English Language Testing and Evaluation (ELTE) in teacher training programs offers LAL training for its pre-service teachers in Turkey. The main syllabus of ELTE course is suggested by The Council of Higher Education (CoHE), which is responsible for all the standards and implementations of higher education in Turkey. Nevertheless, course lecturers or universities are autonomous to design their course contents, which leads the content of ELTE being mostly organized by its course lecturers. It should be noted that each course lecturer has own understanding and style of teaching; and thus, the ELTE content can differ among universities.

Since the current research is carried out in the Spring Semester of 2017-2018 Academic Year, the content of ELTE course is liable to the CoHE program of ELT accepted in 2007. Although there is a new program for ELT teacher education accepted in 2018, it has commenced in 2018-2019 Academic Year. Therefore, the sample studied in this study were not a part of the new program; they were educated under the 2007 program. According to 2007 CoHE program, ELTE is placed in the last semester of

teacher training as a compulsory course and delivered 3 hours a week. The content of ELTE include such topics as (CoHE-ELTE Course, 2007, p. 17):

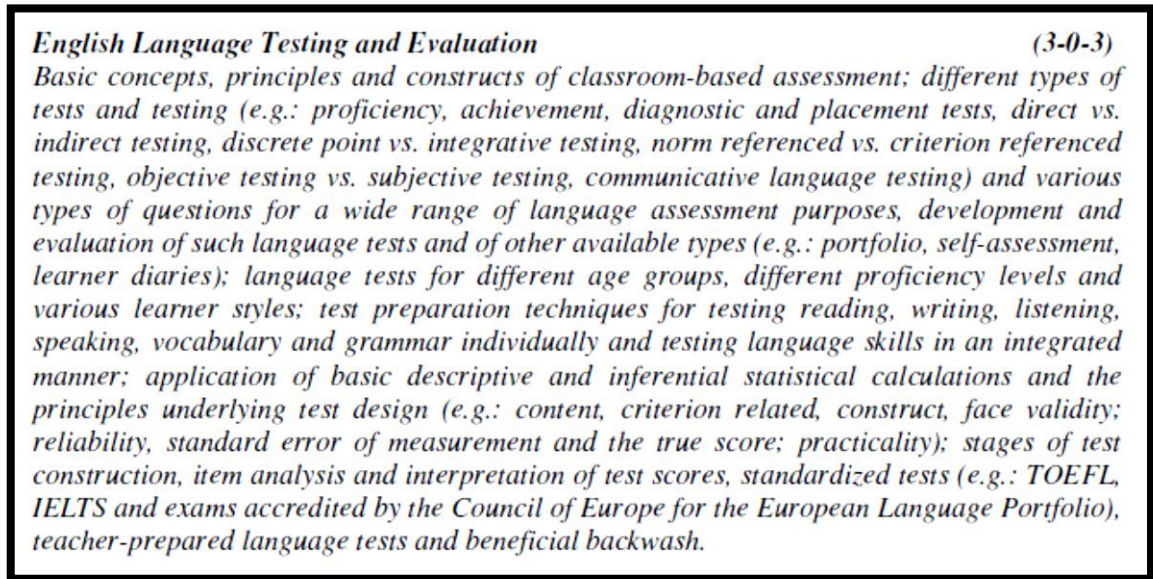


Figure 2.5. The suggested course content for ELTE by CoHE

As can be seen, most of the topics are devoted to English language classroom-assessment for each level and age group of students. Then, pre-service ELT teachers are expected to acquire such knowledge and skills about language assessment before graduation. Besides, most of ELT graduates are appointed by a central institution in Turkey, named Ministry of National Education (MoNE), as EFL teachers to teach English across Turkish different state school settings (primary, elementary, secondary, etc.). Therefore, EFL teachers are expected to have certain competencies before finishing the training program successfully. The competencies of Turkish EFL teachers recommended by MoNE contain the qualifications of planning and organizing the process of ELT; developing students' language skills; monitoring and assessing language development; collaboration with school, family and community; and maintaining professional development (MoNE-ELT Competencies, 2008). Out of such qualifications, the competence of monitoring and assessing language development/progress of students requires EFL teachers to be language assessment literate. In terms of this qualification, EFL teachers must possess the competencies of determining the aims of measurement and evaluation; using instruments and methods of measurement and evaluation; interpreting results of assessment and giving feedback to identify the language

development; using the results of assessment in teaching practices (MoNE-ELT Competencies, 2008). Therefore, the standards established by MoNE can be a guideline for language assessment courses in teacher education. However, it can be inferred that the ELTE course content by CoHE respond to the qualifications of Turkish EFL teachers set by MoNE when compared with each other.

All in all, though there seems to be a standard syllabus for language assessment training suggested by CoHE in Turkey, as mentioned before, the content is shaped by its course lecturers to provide pre-service teachers with required competencies of language assessment; and hence, there might be different practices in terms of ELTE course content.

2.4.3.2. Research studies on LAL training in Turkey

In Turkey, there are few studies conducted with pre-service ELT education with regard to LAL training. Therefore, more research into this topic is needed to reveal how ELTE course is carried out to sustain LAL development.

Hatipoglu (2010, 2015) made a significant contribution to the field by studying the language assessment and testing course given at one university in Ankara, Turkey. Both of her studies were the evaluations of that course in the program.

Her first study (Hatipoglu, 2010) was conducted to demonstrate how a language testing and evaluation course in English language teacher training program was maintained, and what the reflections of undergraduates about that course were. To address those issues, she administered questionnaires ($N=81$) and interviews ($n=16$) with pre-service English language teachers. The results showed that the participants thought the topic of testing skills/knowledge was the most useful topic for their future. Besides, the sample reported that they did not have enough practice about language assessment due to lack of time and student presentations were dominant in the course. Therefore, they stated there must be changes in those issues. The researcher indicated that the ideal syllabus designed by the course instructors was not matched with the needs of students because they were not much meaningful according to the sample.

In her other study (Hatipoglu, 2015), with the aim of preparing a course syllabus, she explored the needs of pre-service English language teachers' assessment training needs. That study lasted for three years, and during that process, she collected data about the needs of language assessment via survey and focus group interviews. A total of 124

pre-service teachers took part in the study. The analysis of the data revealed that previous experiences with testing had an impact on the undergraduates' beliefs of assessment. The sample also mentioned local needs (i.e., Turkish examination system) affected attitudes towards testing and assessment. The insufficiency of one course for LAL training was underlined by the participants, as well. The most rated topics about language assessment were testing-oriented rather than assessment. Then, the researcher concluded that the content of ELTE course should be tailored according to the needs of pre-service ELT teachers.

The last study about LAL training in ELT teacher education was executed by Sariyildiz (2018). She investigated the development of LAL of pre-service ELT teachers. More specifically, she explored their perceptions of theoretical and practical LAL training in teacher education program in line with the emphasis on the ELTE course as well as the practicum, and also, whether further training was needed in that sense. To examine those issues, undergraduate students who were at their last year of the program and were doing their teaching practice at the same time participated in the study. 101 pre-service teachers of ELT responded to a questionnaire of LAL, and 25 teacher candidates in the same program were interviewed about their perceptions of LAL. The findings revealed that the LAL training they took was not enough although they learned how to assess language skills and areas. Thus, they emphasized that they needed training in different domains of assessment, as well. The participants mostly complained about not practicing assessment in the practicum; that is, teaching practice did not cover LAL training. Although the participants appreciated the importance of ELTE course, they stated they needed to put their knowledge into practice, especially by means of teaching practice so that they become competent EFL teachers helping their future students to improve more in the target language by using assessment as a teaching tool.

Here is the summary of the research studies conducted in Turkish context regarding LAL training (see Table 2.6).

Table 2.6. *Research studies conducted in Turkish context about LAL training*

	Author(s)/Year	Context	Topic	Methodology	Results/Findings
1	Hatipoglu (2010)	Turkey – ELT – training	how ELTE course was practiced	- 81 pre-service teachers for questionnaire - 16 pre-service teachers for interview	- most thought topic of testing skills/knowledge was most useful - not have much practice due to lack of time

					- student presentations dominant - syllabus did not meet needs of pre-service teachers
2	Hatipoglu (2015)	Turkey – ELT – training	needs about assessment training via ELTE course	- 3 year-study - 124 pre-service teachers - survey and focus group	- previous assessment experience affected beliefs - Turkish examination system affected attitudes - one course insufficient for training - topics of the course testing based
3	Sariyildiz (2018)	Turkey – ELT – training	perceptions and practices of LAL in teaching practice	- 101 pre-service teachers for questionnaire - 25 pre-service teachers for interview	- training not enough to assess - not practicing assessment in the practicum - aware of the importance of ELTE in teaching

As a final remark, despite few studies about language assessment courses in Turkey, there are significant contributions in terms of assessment training needs of pre-service ELT teachers. More research into this topic will provide substantial insights about LAL training in Turkey.

CHAPTER 3

3. METHODOLOGY

3.1. Type of Research

The current study is conducted to discover language assessment training needs of pre-service English language teachers, and also to find out the effect of ‘English Language Testing and Evaluation (ELTE)’ course on the development of teacher candidates’ LAL, and to reveal whether there is a change after the course is taken. Hence, two aims, needs analysis and evaluation of the course for LAL training, guide the present research. The research type is descriptive in essence because it has the purpose of describing, explaining and exploring the form, structure, activity or change of LAL within an English teacher education program through the eyes of the concerned stakeholders.

Descriptive research, as the name suggests, deals with the description of variables. It aims to portray the features of individuals, situations or groups. In other words, ‘the major purpose of descriptive research is description of the state of affairs as it exists at present’ (Kothari, 2004, p. 2). This type of research is advantageous in terms of two points: It may increase the knowledge of what happens, or how and why it happens regarding the phenomenon, and also, might provide insights by comparing and contrasting to make further implications (Kothari, 2004). In this type of research, researchers can only report what characteristics of variables are in their own natural way without control. In this respect, characteristics, opinions, perceptions, beliefs, attitudes and experiences about a situation are explored through the viewpoint of its participants in order to obtain a detailed understanding of the given situation to describe. Since descriptive research is mainly concerned with characteristics or state of affairs through participants’ views or their natural states, it focuses mostly on qualitative descriptions (Fraenkel, Wallen & Hyun, 2011). These kinds of descriptions can be achieved through observation, survey, interview, document and other sources of qualitative data (Gravetter & Forzano, 2012; Koh & Owen, 2000).

To conclude, considering the purposes of the current research, descriptive research was preferred so as to illustrate the related thoughts and experiences of the stakeholders about LAL to describe, reveal and evaluate assessment training needs within the context of ELT teacher education by means of ELTE course.

3.2. Research Context

In Turkey, there is a four-year English language teacher education program offered by universities in the form of Bachelor of Arts (BA) Degree. The medium of instruction in such a program is mainly English, and so, undergraduate students take compulsory and elective courses of ELT in addition to other general educational and pedagogical courses taught in Turkish for teaching profession. In the first year of the training program, students take language skills and areas courses such as speaking, writing, reading, listening, and grammar. In the proceeding years, they take linguistics, literature, translation, language acquisition, teaching methods, material design, and assessment courses. In their last year, they go to the practicum to do teaching practice. After they complete all their courses successfully, they are graduated and be ready to teach different age groups and levels of English language learners in different school types across Turkey. In terms of assessment and evaluation training, there are two courses at ELT departments: One of them is called ‘Measurement and Assessment in Education’, which is delivered in Turkish language, common to all the departments of the education faculty, and about the general educational assessment concepts, practices, and psychometric evaluations; the other one is called ‘English Language Testing and Evaluation’ which is directly related to the testing of language skills, and the concepts of English language assessment. Thus, it is a subject-matter course complementary to the other course in the program.

The focus of this research is on LAL training in ‘English Language Testing and Evaluation (ELTE)’ course in the undergraduate teacher education program because the effectiveness of this course is explored in order to develop LAL of pre-service English language teachers before they begin teaching. Therefore, LAL training can be said to be dependent on course-based training in the present study. As a representative sample, the course of ELTE given at Anadolu University is investigated by means of the needs, opinions and experiences of its present course-takers and graduates, and also, the investigation is supported by the evaluations of other course lecturers’ opinions and implementations about this course at their own universities and the documents about how this course shaped in the participant course lecturers’ universities.

The ELTE course at Anadolu University is given in the last semester, i.e., the 8th semester, of the teacher training program. It is a compulsory content course instructed by four academicians (in the Spring Semester of 2017-2018 Academic Year). The course

syllabus consists of basic terms and concepts about language assessment, and the design and evaluation of different English language skills testing tools (see Appendix-1 for the syllabus). The main coursebook used in this course is Heaton's (2011) 'Writing English Language Tests' (see Appendix-2 for the book content) and supplied by other testing resources as the coursepack. All the topics are devoted to classroom-based assessment and teacher-made assessment because most graduates of ELT start teaching English in various types of schools and for different age groups of learners. They also use testing and assessment to regulate teaching. Thus, they need to develop their own testing instruments and methods to assess their students in addition to their teaching job rather than designing only standardized-tests which require additional expertise in the area. This course is, then, assumed to give the required knowledge and skills about assessing English as a foreign language in the context of classroom.

3.3. Participants

The present study is designed to have two phases: Needs analysis and evaluation phases to describe LAL training in ELTE course. Therefore, there are three groups of participants (as stakeholders) asked for their opinions and experiences about English language testing and assessment in relation to ELTE course. All the participants are non-native speakers of English in an EFL setting (Turkey) and their first language is Turkish.

In the first group, there are a total of 10 course lecturers who give English language testing and assessment course at the ELT departments of 7 universities (Anadolu $n=4$; Osmangazi $n=1$; Bogazici $n=1$; Middle East Technical $n=1$; Hacettepe $n=1$; Marmara $n=1$; Uludag $n=1$). At first, 12 universities, 21 course lecturers were invited to respond to the questions; however, only 10 lecturers from 7 universities responded. In the second group, there are novice EFL teachers who graduated from Anadolu University within five years; so, the graduates of years of 2013, 2014, 2015, 2016, and 2017 would take part in the study (nearly 800 alumni). The reason why only the graduates of Anadolu University were included in the study is that they had taken ELTE course based on the same syllabus in recent years carried out for a long time by the same academicians as in pre-service teachers. However, 37 teachers filled the forms of open-ended questions about the mentioned course; 15 of them were excluded due to their school type (university or private school) causing inconsistency about the study aims. It is because these types of schools have their own testing offices working with a team of experts. Therefore, the rest,

22 teachers from primary, elementary and high schools in Turkey, was the main group of participators as novice EFL teachers. The last group contains a representative sample of pre-service English language teachers who took ELTE course in the Spring Semester of Academic Year 2017-2018 at Anadolu University. They were asked for their opinions two times: Before taking the course and after taking the course. Therefore, at the beginning of the semester, 164 pre-service ELT teachers responded to open-ended questions and at the end of the semester, 146 course-takers participated. All the participants took part in the research through purposeful sampling on a voluntary basis. See Table 3.1 for the detailed distribution of the sample.

Table 3.1. *The distribution of the participants (numbers)*

	<i>Invited</i>	<i>Responded</i>	<i>Excluded</i>	<i>Rest (main one)</i>
Universities/ Course Lecturers	Anadolu = 4	Anadolu = 4		<i>total:</i>
	Osmangazi = 1	Osmangazi = 1		
	Bogazici = 4	Bogazici = 1		7 universities
	METU = 2	METU = 1		10 lecturers
	Hacettepe = 1	Hacettepe = 1		
	Gazi = 1	Marmara = 1		
	Istanbul = 1	Uludag = 1		
	Marmara = 1			
	Yildiz Teknik = 2		N/A	
	Dokuz Eylul = 1			
	Akdeniz = 1			
	Uludag = 2			
	<i>total:</i> 21 lecturers, 12 universities	<i>total:</i> 7 universities, 10 lecturers		
Novice EFL teachers (Alumni)	nearly 800 (between 2013 and 2017)	37 (both state and private schools)	15 (university/ private schools)	<i>total = 22</i> state: primary = 5 elementary = 7 high school = 10
Pre-service ELT teachers (course- takers) -BEFORE-	nearly 200 (2017-2018 Spring Semester)	168	4	164
Pre-service ELT teachers (course- takers) -AFTER-	nearly 200 (2017-2018 Spring Semester)	147	1	146

As far as the characteristics of each sample in terms of English language assessment (ELA) are considered, the instructors evaluated themselves as highly competent in the field of English language testing and assessment (see Table 3.2).

Table 3.2. *The profile of the course lecturers*

<i>Name</i>	<i>Years of teaching experience</i>	<i>Perceived level in ELA</i>
Anadolu.1	15	high
Anadolu.2	10	high
Anadolu.3	15	high
Anadolu.4	15	high
Osmangazi.1	3	high
Bogazici.1	4	high
METU.1	2	high
Hacettepe.1	12	high
Marmara.1	18	high
Uludag.1	14	high

Besides, even though most of the graduates perceived themselves highly competent in English language assessment (ELA), there have still been some who thought their skills of testing and assessment were poor and needed improvement, especially the teachers of higher-level language learners (see Table 3.3).

Table 3.3. *The profile of the novice EFL teachers (the alumni) (numbers)*

<i>Graduation year</i>	<i>School grade (all state)</i>			
	<i>primary</i>	<i>elementary</i>	<i>high school</i>	<i>total</i>
2017	0	2	1	3
2016	1	3	3	7
2015	2	1	2	5
2014	1	1	2	4
2013	1	0	2	3
<i>total</i>	5	7	10	
<i>Perceived level in ELA</i>				
high	2	3	6	11
moderate	2	4	3	9
low	1	0	1	2
<i>total</i>	5	7	10	

As for pre-service ELT teachers, most of them felt themselves competent in English language assessment (ELA) at the average level ($n=109$) after they took ELTE course at the end of the semester (see Table 3.4).

Table 3.4. *The perceived levels of ELA by the pre-service teachers after the course (numbers)*

<i>Distribution of levels</i>	<i>n</i>
High	33
Moderate	109
Low	4
<i>total</i>	<i>146</i>

3.4. Data Collection Procedure

The current study addresses two research questions about language assessment literacy. Accordingly, open-ended questions for each stakeholder of ELTE course were formed, and the documents regarding webpages of participant universities about the course were collected in the first place. Thus, there are two types of qualitative data sources. Then, the data by means of the mentioned materials were gathered in three steps. All the collection procedures were executed by the researcher herself.

As the data collection tool, open-ended questions were preferred instead of closed-ended surveys because they provide a deeper understanding of the topic at hand by displaying the causes, reasons, experiences, perceptions, views and the like behind the inquired concept without giving predefined categories (Creswell, 2012). Besides, instead of interviews, open-ended questions are more useful to reach a larger sample (Gay, Mills & Airaisan, 2012). All the open-ended questions for each group of participants (ELTE course lecturers, novice EFL teachers-alumni, pre-service ELT teachers) were constructed by the researcher herself in line with the relevant research studies and papers in the literature. The questions were written in Turkish because the participants are believed to express themselves better in their mother tongue. For the purposes of reliability and validity, three experts in the field reviewed the questions. After this kind of expert checking, different numbers of questions were obtained: 10 questions for ELTE course lecturers (see Appendix-3); 6 questions (at the beginning of the semester) and 6 questions (at the end of the semester) for pre-service ELT teachers (see Appendix-4); 8 questions for the alumni (see Appendix-5). The consent forms for each group of participants were also added owing to ethical concerns (see Appendix-6). The informed consents were also used to show voluntariness of the sample, the privacy, confidentiality and anonymity of the data in order to indicate research integrity. In addition, a categorical checklist was designed as comprehensive as possible by the researcher herself to analyze the documents by looking at the categories given in the information part of the webpages about ELTE course of the participant universities (see Appendix-7).

Before the data collection commenced, certain preparations were made. First, the application to the institutional review board for the research ethics approval of the present study was submitted, and the permission was granted (see Appendix-8). Second, open-ended questions accompanied by consent forms were generated by means of 'Google Forms' as the online platform to reach the sample (course lecturers and novice EFL teachers) (see Appendix-9 for the samples). The questions were also printed out for pre-service ELT teachers. Third, the e-mail addresses of course lecturers were found through their websites in the related university pages. Fourth, the e-mail addresses of the alumni, novice EFL teachers, were granted by the courtesy of the registrar's office at Faculty of Education, at Anadolu University. Besides, the alumni were contacted via their previous consultants at Anadolu University through Facebook. Last, the application time of open-ended questions to pre-service teachers was organized by means of getting permissions from the course lecturers at Anadolu University.

After such sort of preparations, the data collection began in three steps.

In the first step, open-ended questions both for ELTE course lecturers and novice EFL teachers were sent via the Internet platform (Google Forms) to their e-mails. Firstly, ELTE course lecturers of 12 universities (21 instructors) were contacted and invited to participate in the study. However, 10 instructors (7 universities) responded to the questions. Then, the alumni as novice EFL teachers (nearly 800 graduates of ELT program of Anadolu University) were invited to answer the questions and share their opinions via e-mail to the 'Google Forms'. They were also invited through Facebook by the help of their former consultants at university. Only 37 novice EFL teachers responded to the questions. But, 15 of them were excluded due to their inconsistency with study aims; so, 22 teachers left as the main participants. The responses given in 'Google Forms' were saved in Microsoft Excel document format automatically by the mentioned online platform (see Appendix-10 for the samples). This period lasted from the end of Fall Semester (December 2017) to the end of Spring Semester (June 2018) of 2017-2018 Academic Year.

In the second step, open-ended questions accompanied by consent forms for pre-service ELT teachers were given in the form of paper at their regular class meetings (see Appendix-11 for the sample). The data about them were collected two times: before taking the ELTE course (February 2018), and after the course (May 2018). At the beginning of the course semester, 168 course-takers responded to questions, but 4 of them

were eliminated because their papers were not filled. Thus, there were 164 participants for the first time. At the end of the course semester, 147 undergraduates filled the forms. However, 1 of them was excluded due to the incomplete paper. Then, 146 participants took part for the second time. This period covered the Spring Semester of 2017-2018 Academic Year.

In the last step, the webpages of the participant universities were examined, and the pages about ELTE course were saved as screenshots for the forthcoming document analysis (see Appendix-12 for the samples). The webpages are official documents but are open to public (Bogdan & Biklen, 2007; Gay, Mills & Airasian, 2012). The documents as another data source contribute to the research in terms of data triangulation (Bowen, 2009) since the analysis of the webpages is believed to verify the same information maintained by the participants as well as strengthen the credibility of the study.

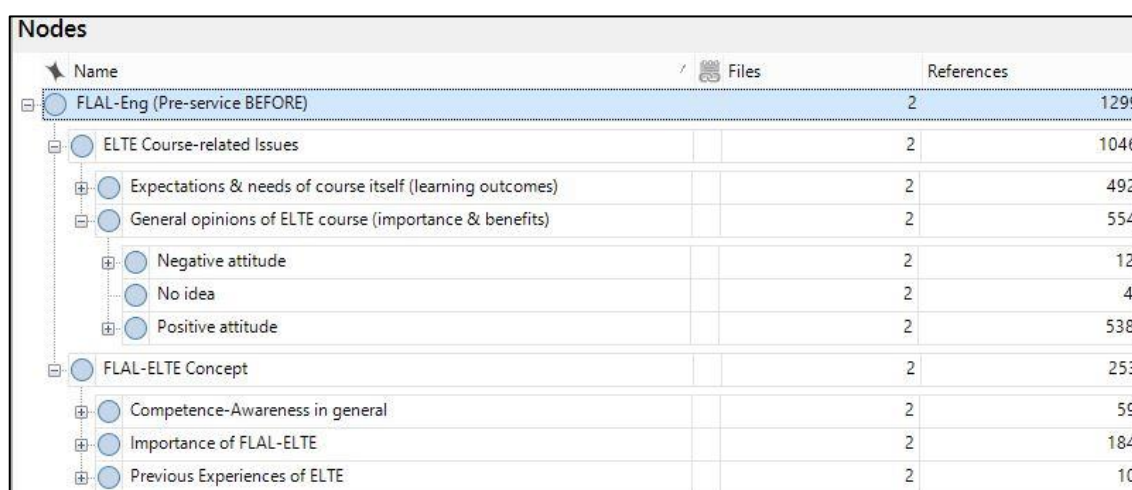
In conclusion, the reliability and validity of the present study was tried to be established through the detailed descriptions of the research methodology (participants, instrument, collection, analysis), contributions of field experts, collecting the evidences of different types of data sources (open-ended questions and documents) from different participant groups purposefully, and using intercoder agreement for the analysis (Creswell, 2012; Miles, Huberman & Saldana, 2014; Patton, 2002; Yin, 2011).

3.5. Data Analysis

The present study is based on the descriptive research and utilizes open-ended questions and documents to collect the data. Hence, the process of data analysis was dominantly qualitative; namely, inductive and interpretive. According to the data, there were two types of analysis: content analysis for the open-ended questions, and document analysis for the webpages.

Since the responses to open-ended questions provided qualitative data, content analysis was carried out. Content analysis refers to ‘searching text for recurring words or patterns.....any qualitative data reduction and sense-making effort that takes a volume of qualitative material and attempts to identify core consistencies and meanings’ (Patton, 2002, p. 453). For content analysis, Creswell’s (2012), Miles and Huberman’s (1994), and Miles, Huberman and Saldana’s (2014) recommendations were taken into consideration. First of all, the written responses of pre-service ELT teachers were transcribed into Microsoft Word document by the researcher herself. Then, the online

gathered responses of ELTE course lecturers and novice EFL teachers were transferred into Microsoft Word document by the researcher from Microsoft Excel document that was automatically constructed by ‘Google Forms’. All the Word documents were revised and edited in terms of their spelling, punctuation and the like. Afterwards, all the data for each stakeholder were examined separately by hand as the preliminary exploratory analysis in order to get the general sense of data. Few chief codes and themes were determined, which was conducted inductively (see Appendix-13 for the versions by hand). That first exploration was shown to the advisor as the expert to get feedback, and then, the data were embedded into NVivo 11 Pro program for a detailed content analysis. While the codes were being labeled, essential major and minor themes were determined at the same time. The number of communication units or data chunks found via codes was also automatically computed under the title of ‘references’ in the program and shown in frequencies. All the codes were revised by means of reducing overlapping and redundant ones, and after that, all of them were collapsed into the relevant themes meanwhile the categorical themes were interrelated. At the end of such organization of codes and themes in the program, the list of the findings for each stakeholder was generated for expert checking due to reliability and validity issues. Two ELTE course lecturers at Anadolu University evaluated the findings. Some misunderstandings in the codes were refined in order to make clearer meaningful wording, and the names of some subthemes were discussed. After such feedback session, code-theme relationship was established. The last versions of the code-theme organization can be seen in Appendix-14 and here is one of the sample screenshot for the organization of themes:



Name	Files	References
FLAL-Eng (Pre-service BEFORE)	2	1299
ELTE Course-related Issues	2	1046
Expectations & needs of course itself (learning outcomes)	2	492
General opinions of ELTE course (importance & benefits)	2	554
Negative attitude	2	12
No idea	2	4
Positive attitude	2	538
FLAL-ELTE Concept	2	253
Competence-Awareness in general	2	59
Importance of FLAL-ELTE	2	184
Previous Experiences of ELTE	2	10

Figure 3.1. The sample screenshot for the categorical themes

The interrater-reliability procedure was also used to establish the reliability of the data analysis. For this, Miles and Huberman's (1994) intercoder agreement procedure was applied between two expert controllers and the researcher. In this procedure, the number of agreements and disagreements are compared (p. 64). As a result of that calculation, high percentages of agreement were found between the coders ranging from 95% to 98% for each analysis (findings of course lecturers, novice teachers, and pre-service teachers). In total, 97% of agreement was achieved, which shows there is a high consistency between the raters.

As far as the webpages about ELTE course of participant universities are considered, document analysis was used because it is helpful to review and evaluate documents systematically (Bowen, 2009). This analysis was conducted with the help of the aforementioned researcher-made checklist, which have pre-assigned categories to examine the documents constructed as a result of similar titles found in the documents. After the documents were analyzed, the similarities and differences were found so as to enlighten the issue. Thus, this analysis was believed to be beneficial to evaluate ELTE courses of universities in addition to the analysis of responses provided by the sample. That is, documents were used to support the findings, explain and describe the research context by means of providing the background knowledge for the content of ELTE course. The findings of the document analysis were tabulated and added to discussion when needed.

Finally, all the data findings were evaluated together to get a whole meaning of the concerned topic. The interpretations as a result of making sense of the data were given in the form of narrative discussion (Creswell, 2012) accompanied by illustrating tables and figures, and the quotations from the responses and documents. While discussing the findings, the exemplifying quotes extracted from the responses of the sample were translated from Turkish to English by the researcher herself. However, to eliminate the translation errors and meaning loss, the translated quotes were checked by another ELT researcher. She compared Turkish and English versions, and corrected them to prevent ambiguity in the quotations. All the Turkish-English versions of quotations can be seen in Appendix-20. For the sake of anonymity, certain acronyms for the owners of quotations were used systematically. To exemplify them, see Table 3.5:

Table 3.5. *Acronyms used for the owners of quotes*

CL (Course Lecturer)	A.CL.1: Anadolu Course Lecturer-1 B.CL.5: Bogazici Course Lecturer-5 O.CL.6: METU Course Lecturer-6 H.CL.7: Hacettepe Course Lecturer-7		E.CL.8: Osmangazi Course Lecturer-8 M.CL.9: Marmara Course Lecturer-9 U.CL.10: Uludag Course Lecturer-10	
Level.T (Novice EFL Teachers as Alumni)	Level: perceived level in FLAL (high/moderate/low)			
	High.T.1	Moderate.T.2	Low.T.3	
Before.Pre (Pre-service ELT teachers before course)	Before.Pre.1	Before.Pre.2		
Level.After.Pre (Pre-service ELT teachers after course)	Level: perceived level in FLAL (high/moderate/low)			
	High.After.Pre.1	Moderate.After.Pre.2	Low.After.Pre.3	

CHAPTER 4

4. FINDINGS AND DISCUSSION

In this section, firstly, the findings from each stakeholder and the documents about ELTE course content were presented. Secondly, the summary of the findings and the summary of the responses to each research question were given. Lastly, the discussion about the findings was shared.

4.1. Findings from the Participant Stakeholders and the Documents

The stakeholders are the participants as the course lecturers, novice EFL teachers, and pre-service ELT teachers who shared their opinions and experiences about ELTE course through open-ended questions; namely, LAL training within the scope of ELT training program. The documents about the course content, instructed at the universities where participant course lecturers work, were also collected for further information and details. All the collected data were analyzed, and crucial findings were revealed.

4.1.1. Findings from course lecturers

The course lecturers teaching ELTE in different universities were asked to share their opinions, perceptions and experiences about foreign language assessment and its training at ELT departments in teacher education programs. A total of 10 lecturers from 7 universities responded to the open-ended questions of FLAL in relation to ELTE course. The responses indicated that course lecturers mostly focused on certain keywords while discussing their opinions and experiences:

‘Testing, measuring, theoretical knowledge, experience, language proficiency, ELTE course content, topics and materials, competencies, ability, changes in terms of awareness, language skills and areas, practice, update, willingness, and decision-making’

In relation to these, the data analysis produced basically two themes for the course lecturers: *FLAL-ELTE Concept* (Foreign Language Assessment Literacy-English Language Testing and Evaluation Concept), and *ELTE Course-related Issues* (see Figure 4.1). The course lecturers commented more on ELTE course-related issues ($f=202$) than FLAL-ELTE concept ($f=93$) (see Appendix-15 for detailed tabulation).

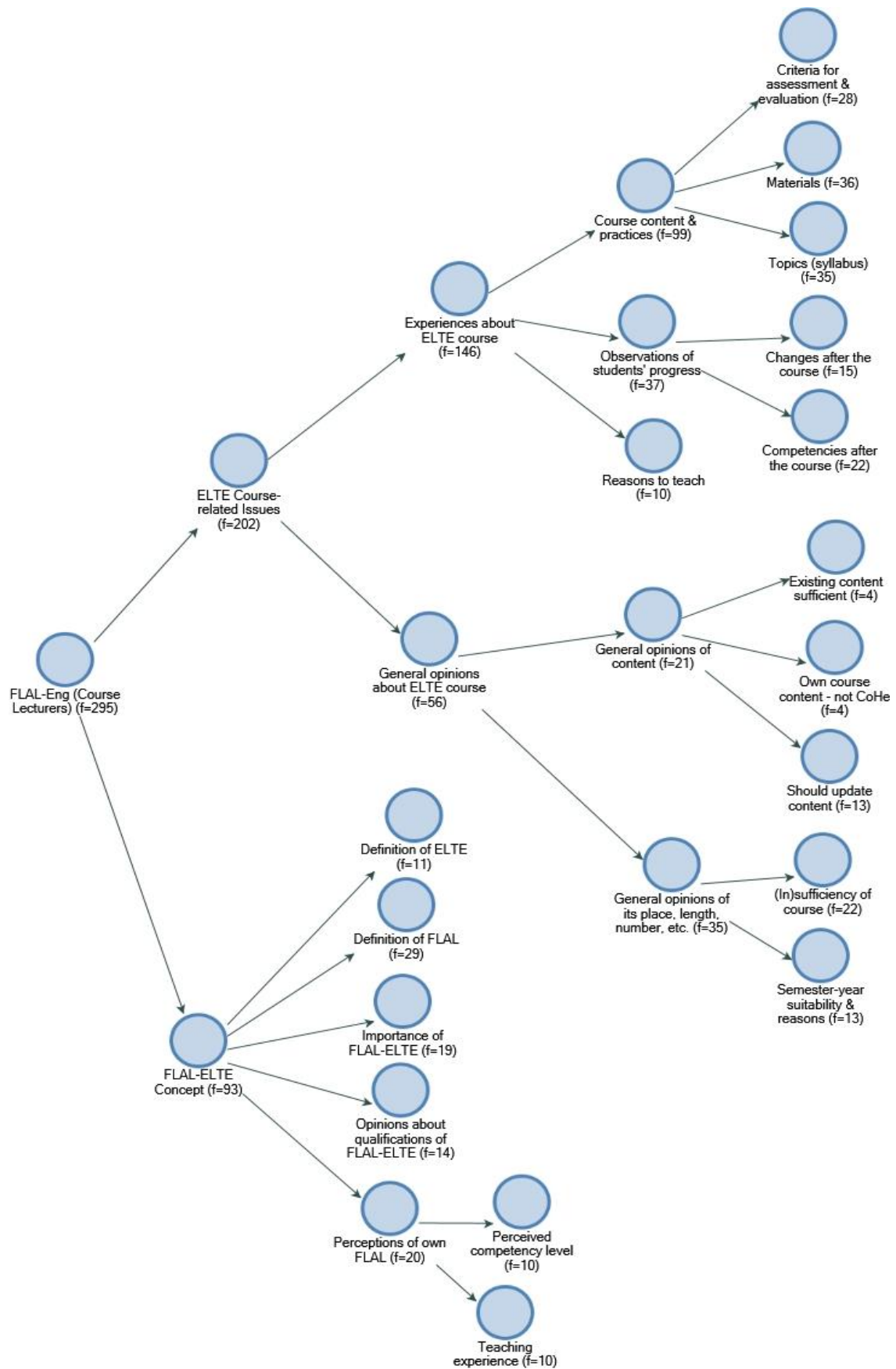


Figure 4.1. *The thematic display of the findings from the course lecturers*

For *FLAL-ELTE Concept*, there are four subcategories: They gave definitions of FLAL ($f=29$) and ELTE ($f=11$), mentioned the importance of FLAL-ELTE itself ($f=19$), shared their opinions about the qualifications of FLAL-ELTE ($f=14$), and reported their perceptions of their own FLAL ($f=20$) (see Figure 4.2).

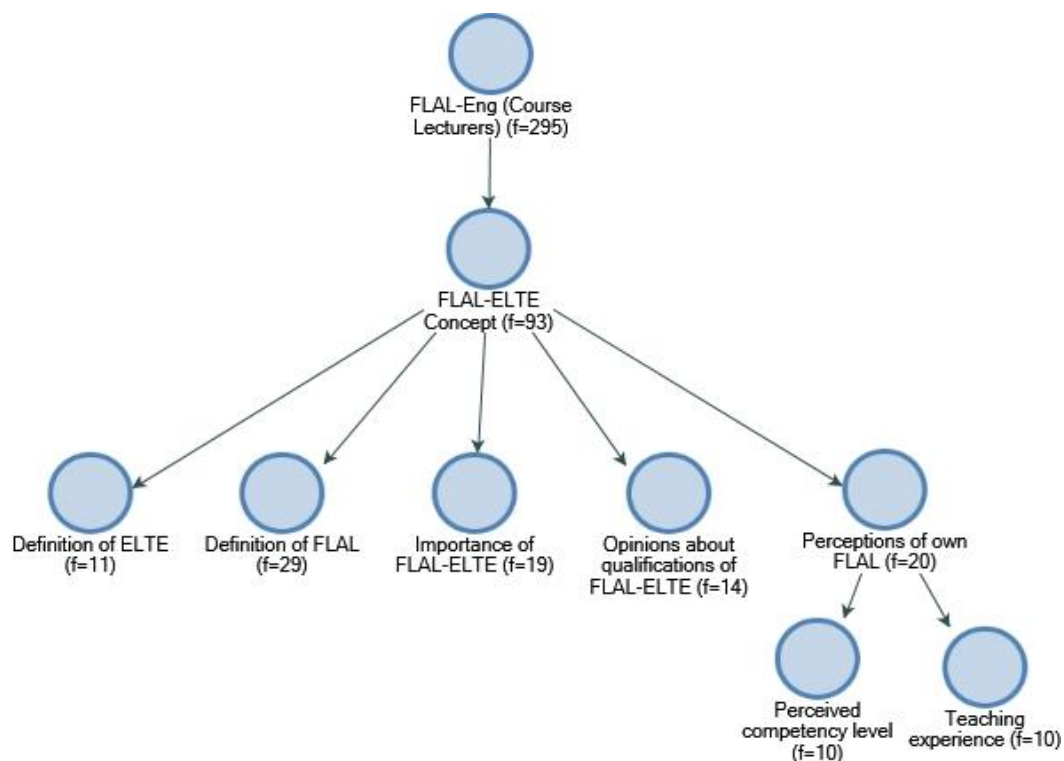


Figure 4.2. *The thematic display of the theme ‘FLAL-ELTE Concept’ for course lecturers*

To begin with, course lecturers defined ELTE as measuring language proficiency of learners (their knowledge and skills in the target language) according to the course objectives; that is, whether learning outcomes of language learning are achieved at the expected level. For example, one of them defined:

‘Testing and evaluation is assessing language proficiency of a foreign language learner on the basis of its use with respect to learning purposes.’ (O.CL.6)

They also mentioned that testing and evaluation are used to determine students’ language level and their deficiencies in that language in order to make decisions about their language progress by using different kinds of assessment procedures. One of them indicated that:

‘Testing and evaluation is assessing effectively language proficiency as carrying out both summative and formative assessment by means of reliable and valid testing tools.’ (E.CL.8)

As a whole, they pointed out that assessment is used to evaluate both learning and teaching process itself. For instance, a course lecturer expressed that:

‘ELTE can be defined as assessing whether language skills and areas in a foreign language are learned or not, and evaluating the teaching process from the perspective of teachers themselves, students and process itself.’ (A.CL.4)

With regard to the definitions of FLAL, course lecturers highlighted different competencies of being assessment literate in English language. They reported that the understanding of concepts and principles of English language assessment, the need to have basic methods, knowledge and skills of ELTE, and the ability to develop appropriate testing tools, to practice assessment procedures and to evaluate and interpret test scores so as to make decisions about the process are the essential elements of being FLAL. For instance, one of the course lecturers exemplified such that:

‘The concept of FLAL has actually subcategories. First, you need to have the knowledge about how to test language proficiency. Then, you need to have the ability to design language tests and the competence to interpret assessment results appropriately in the light of statistical knowledge. Besides, you need to pay attention certain principles and rules while preparing such tests. For example, there is a certain way of teaching style used in your course, and accordingly, you have to develop an evaluation system in which you can prepare related questions, construct evaluation criteria and obtain reliable and valid findings.’ (A.CL.3)

Besides, one of the participant course lecturers underscored the literacy issue by stating that language assessment literacy refers to being expertise in all the aspects of foreign language assessment. Her comment is like that:

‘Literacy can refer to the expertise in any profession or application which needs to be knowledgeable about its requirements to be able to fulfill them. In terms of language concept, I can define literacy as having the expertise in an appropriate assessment procedure through which foreign language learners’ proficiency is assessed with respect to the aims.’ (B.CL.5)

On the other hand, only one course lecturer drew attention to the concern of ethics with regard to being language assessment literate. She pointed out that:

‘FLAL is to know when and why assessment knowledge is needed, how to perform effective testing and evaluation, how to interpret and store the results, and how to share the results ethically.’ (U.CL.10)

In addition to these, course lecturers emphasized the importance of language assessment and testing in English language as well as the importance of being assessment

literate in English for better education. They pointed out that assessment is as important as teaching itself because it is a crucial feedback tool for both students and teachers themselves. In other words, not only teachers have the opportunity to assess whether their instruction is effective and to make decisions accordingly, but also students' language level and deficiencies are determined by means of language testing and evaluation. For instance, one of them mentioned that:

‘Assessment is very important. Testing-evaluation is an inseparable concept. If we did not assess properly, we would not be able to take new steps in teaching and to prepare effective plans. The feedback about the main deficiencies of students in English is provided most efficiently and quickly by tests and assignments.’ (A.CL.3)

However, one of the course lecturers underlined the fact that the concept of assessment is not only important in English language teaching, but it is also important in other fields because it is a prominent feedback tool. He commented that:

‘As in other fields [apart from ELT], the purpose of assessment is to give feedback about students, teachers and program itself.’ (H.CL.7)

Another point focused on by the course lecturers under FLAL-ELTE concept theme is that being foreign language assessment literate requires a distinct qualification. They stated that this qualification is a matter of being knowledgeable about the subject and content of ELTE as well as being experienced in assessment. One of them discussed that:

‘Foreign language assessment needs expertise in this field. It is a technical subject in terms of both theoretically and practically.’ (M.CL.9)

However, only one course lecturer concentrated on another point of the qualification in language assessment literacy. She believed that apart from basic knowledge and skills, language testing and evaluation necessitate to know statistics in order to interpret assessment results by communicating with them to give feedback. She argued that:

‘Foreign language assessment requires a specific qualification in this area. In addition to the basic concepts and principles of testing, it is important to know especially statistics. It is not a branch to be learned by reading a book.’ (B.CL.5)

Considering the perceptions of FLAL, all the course lecturers perceived themselves to be highly competent in language assessment ($f=10$) whereas most of them had from 10 to 15 years of teaching experience in ELTE course instruction ($f=6$), which shows they

believed they have the necessary knowledge and skills as well as the experience in English language assessment.

As far as the theme titled *ELTE course-related issues* is taken into account, they focused on their experiences about their own ELTE course ($f=146$) and expressed their general opinions about ELTE course ($f=56$) under two subthemes.

The first of the subcategories is about the experiences regarding their own ELTE course in which they reported how they ran their course ($f=99$), what they observed about their students' progress after the course ($f=37$), and why they chose to teach this subject ($f=10$) (see Figure 4.3).

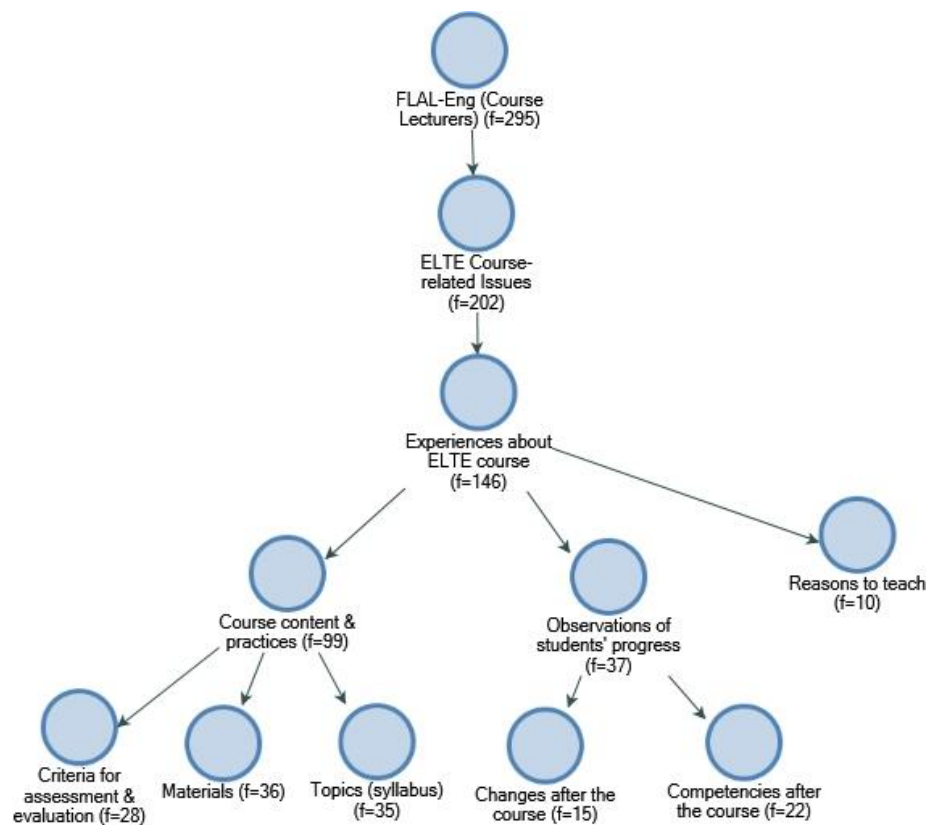


Figure 4.3. The thematic display of the subtheme 'Experiences about ELTE Course' for course lecturers

Pertaining to ELTE course content and practices, course lecturers shared their syllabus in which the topics of the course were listed. While explaining their content titles, they introduced their already existing topics and later added topics in the syllabus in line with the book they used as the coursepack. The existing topics mainly involved basic knowledge and methods of ELTE, types of test items and questions, and assessment of language skills and areas. To illustrate them, one noted that:

‘My syllabus includes the topics of what is testing, what is assessment, principles of classroom testing (reliability, validity, etc.), standardized testing, testing grammar/vocabulary, assessing listening, assessing reading, assessing writing, assessing speaking, integrated testing and assessment.’ (E.CL.8)

In terms of later added topics, even though few course lecturers said they did not add any new topic, most of them emphasized recent approaches and practices in language assessment as their new topics. They mentioned alternative assessment, dynamic assessment, computer-based testing, and so on. For example, one of them stated that:

‘I added the topics of preparing language tests practically and alternative assessment. Besides, I focus mostly on integrated testing because I want students to understand the importance of integrated testing while assessing the overall structure of language itself and language proficiency.’ (E.CL.8)

As for the materials the course lecturers chose to use, in addition to the well-known testing books, most of them added other materials such as homework exercises, articles, sample problems and tasks, and information about statistics. For instance, one course lecturer stressed that:

‘We added sample tests for various age groups and purposes because we wanted students to have a good coursepack if they needed.’ (A.CL.3)

Among the testing books, the most used one belongs to Heaton’s ‘Writing Language English Tests’, and then, Hughes’ book titled ‘Testing for Language Teachers’ comes. They preferred to use those books because of their covering many practical samples in relation to their course aims and students’ needs. For instance, one commented on that:

‘[We chose this book [Heaton’s book] as the coursebook because] it consists of practical examples, and thus, it does not only provide theoretical information.’ (A.CL.2)

Regarding the criteria for assessment and evaluation of their ELTE course, all of the course lecturers mentioned they administered a midterm and a final exam to their students mostly in the form of multiple-choice exams. Apart from those exams, few of them gave other tasks such as presentation, reflection, sample test development exercises, and so forth during the course term. To report the criteria, one of the course lecturers stated that:

‘[As for the assessment and evaluation criteria of ELTE course], students must enter a theoretical test as the midterm exam and prepare a test that measures language skills as the final exam.’ (A.CL.1)

As far as the observations of students' progress throughout ELTE course are considered, the course lecturers discussed the observations in two respects: Changes seen after the course, and competencies that seemed to be acquired after course. In the matter of changes, most of them highlighted the issue of awareness. They argued that students became more aware of ELTE while gaining general knowledge and perspective of language assessment as well as criticizing assessment procedures. Hence, there seemed to be a positive change as cited by the course lecturers. To exemplify;

‘Definitely, there is a change at the end of the course. Students become more aware of assessment and are enlightened about exam preparation and administration. They criticize each exam they take and give importance to each exam they prepare.’ (O.CL.6)

Furthermore, one of the course lecturers drew attention to the concept of evolution in which students began to understand the relationship between teaching and testing. She explained that:

‘My general observation about this course given at the fourth year of teacher training is that students’ teaching methods that they have learned evolve into assessment methods. [...] They are also surprised by the fact that assessment is actually a meticulous job.’ (B.CL.5)

Concerning the competencies that students seemed to gain after they took ELTE course, the course lecturers indicated that they felt more competent in language assessment because they acquired the ability to design test items more properly and learned basic things about testing and assessment in English. For instance, one of them noted that:

‘[After taking this course], students are able to choose more appropriate test items while doing their assignments, and also, they can make necessary adaptations. They are able to write test instructions more properly and pay attention to time allocations among test tasks. Moreover, they can find appropriate evaluation criteria or take into account certain points while preparing such criteria.’ (A.CL.3)

Nonetheless, most of the course lecturers admitted that students were lack of experience and practice of language assessment though they learned theoretical knowledge well. Thus, to illustrate the shortfall of their course, they underscored the problem of lacking practical skills. One of the course lecturers noted that:

‘There is a lack of experience [about language assessment], but at the end of the course, students’ theoretical knowledge is quite good and are more ready to assess.’ (H.CL.7)

When it comes to the reason why they preferred to instruct the ELTE course, most emphasized their willingness and interest in language assessment. For example, one mentioned that:

‘Because of my interest in testing and evaluation, I wanted to teach this course.’ (A.CL.2)

The second of the subcategories of ‘ELTE Course-related Issues’ is about course lecturers’ general opinions of the content ($f=21$) and general opinions of the course’s place, length, number, etc. in the program ($f=35$) (see Figure 4.4).

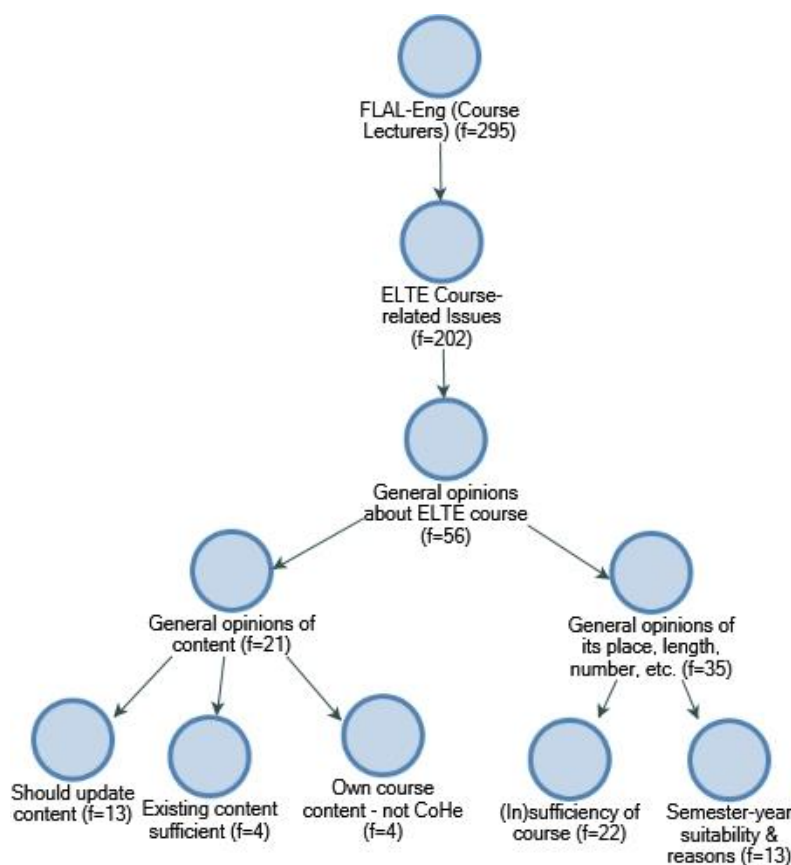


Figure 4.4. *The thematic display of the subtheme ‘General Opinions about ELTE Course’ for course lecturers*

In view of course lecturers’ general opinions about the course content, some of them stated that they were not aware of CoHE’s suggested ELTE course content and organized their own syllabus and selected relevant materials to utilize on their own course. To illustrate, one of them stated that:

‘I have not heard about CoHE’s program [ELTE course content]. We ourselves [course lecturers] determined the topics. We chose appropriate coursebook and are studying it.’ (A.CL.1)

On the other hand, although few of them acknowledged the existing content was sufficient, most disagreed with this opinion because of the need to revise the content. Thus, they tended to update their content. In this respect, they discussed to add such topics as assessing young learners, alternative methods of assessment, techniques of test preparation for specific groups of learners. For example, one indicated that:

‘The topics of process-oriented alternative assessment (use of observation checklists, learning journals and logs, portfolios, progress reports, etc.) should be more emphasized in this course.’ (A.CL.4)

As for the (in)sufficiency of the course itself, course lecturers shared different ideas. Some of them agreed that only one course about ELTE is enough in the program, others complained about the insufficient sides of the course with regard to its being only one course in the curriculum of ELT. Those who supported that only one course is sufficient explained that there are two courses about assessment in ELT program at university: general education one and English-based one. Therefore, they believed these two courses were adequate. One of them reported that:

‘This course [ELTE] is sufficient. There is also a course that is taken in Turkish in the third year of the teacher training program. If these two courses (Turkish general one and ELTE) are learned thoroughly, it will be sufficient in terms of language assessment.’ (A.CL.1)

However, those who believed that only one course is not sufficient for students to gain required competencies of ELTE mentioned a number of reasons such as shortage of time, intensive content, too many topics to discuss and lack of practical skills that caused insufficiency. To improve the course effect, some suggested that assessment may be integrated into the practicum or elective courses might be given. For instance;

‘[Apart from the course], it would be much better in the practicum if students prepared a real English exam, assessed their students and evaluated the results in the lessons they taught.’ (A.CL.3)

Moreover, some of them recommended that there should be two courses devoted to both theory and practice of language assessment. For example, one of them advised that:

‘In fact, only one course [about ELTE] is not enough; there may be another course which is based on putting knowledge into practice and is more alternative assessment oriented in another semester. Students take a general testing and evaluation course in their mother tongue [Turkish], but they cannot achieve meaningful learning and are not able to transfer what they learn in that course to this ELTE course.’ (A.CL.4)

Besides, one of the course lecturers emphasized the fact that all teaching-methods-courses in the program naturally have the element of assessment. Therefore, she advised that:

‘The dimension of assessment is actually embedded in all teaching methods courses. Although learning-oriented and task-oriented assessment can be distinct, the element of assessment is given consciously or unconsciously in such courses [teaching methodology]. In my opinion, after fundamental concepts and principles of testing are learned, assessment might be added as a dimension to methods courses. Only one course for ELTE is inadequate all the time.’ (B.CL.5)

Considering the given semester time suitability of ELTE course, most of the course lecturers agreed on the idea that the current place of ELTE course (the 8th semester) in the ELT curriculum is good because teaching methodology courses just finish, and it may be embedded in the practicum period. One of the course lecturers stressed that:

‘The second term of the fourth year [8th semester] seems quite good due to the suitability of performing practical applications as well as creating awareness.’ (E.CL.8)

Nevertheless, some of them disagreed with the common idea of the suitable place of ELTE course. They discussed that teaching and learning ELTE should begin earlier. One pointed out that:

‘I think if language assessment is taken as a dimension of teaching methods courses, ELTE education may start in the third year.’ (B.CL.5)

As a conclusion, it can be said that the course lecturers more or less held the same opinions towards FLAL-ELTE concept whereas they had somewhat different experiences, practices and opinions about the course itself.

4.1.2. Findings from novice EFL teachers

The novice EFL teachers, the graduates of ELT at Anadolu University within five years, conveyed their thoughts and experiences about English language assessment in terms of both their ELTE training at university and their ELTE applications in their current classrooms. 22 teachers working with different age groups and levels across Turkish state schools took part in the study. Their responses to open-ended questions were analyzed and two themes were found: *FLAL-ELTE Concept* and *Experiences about ELTE (past & present)* (see Figure 4.5). The sample commented more on experiences about

ELTE (past & present) ($f=245$) than FLAL-ELTE concept ($f=115$) (see Appendix-16 for detailed tabulation).

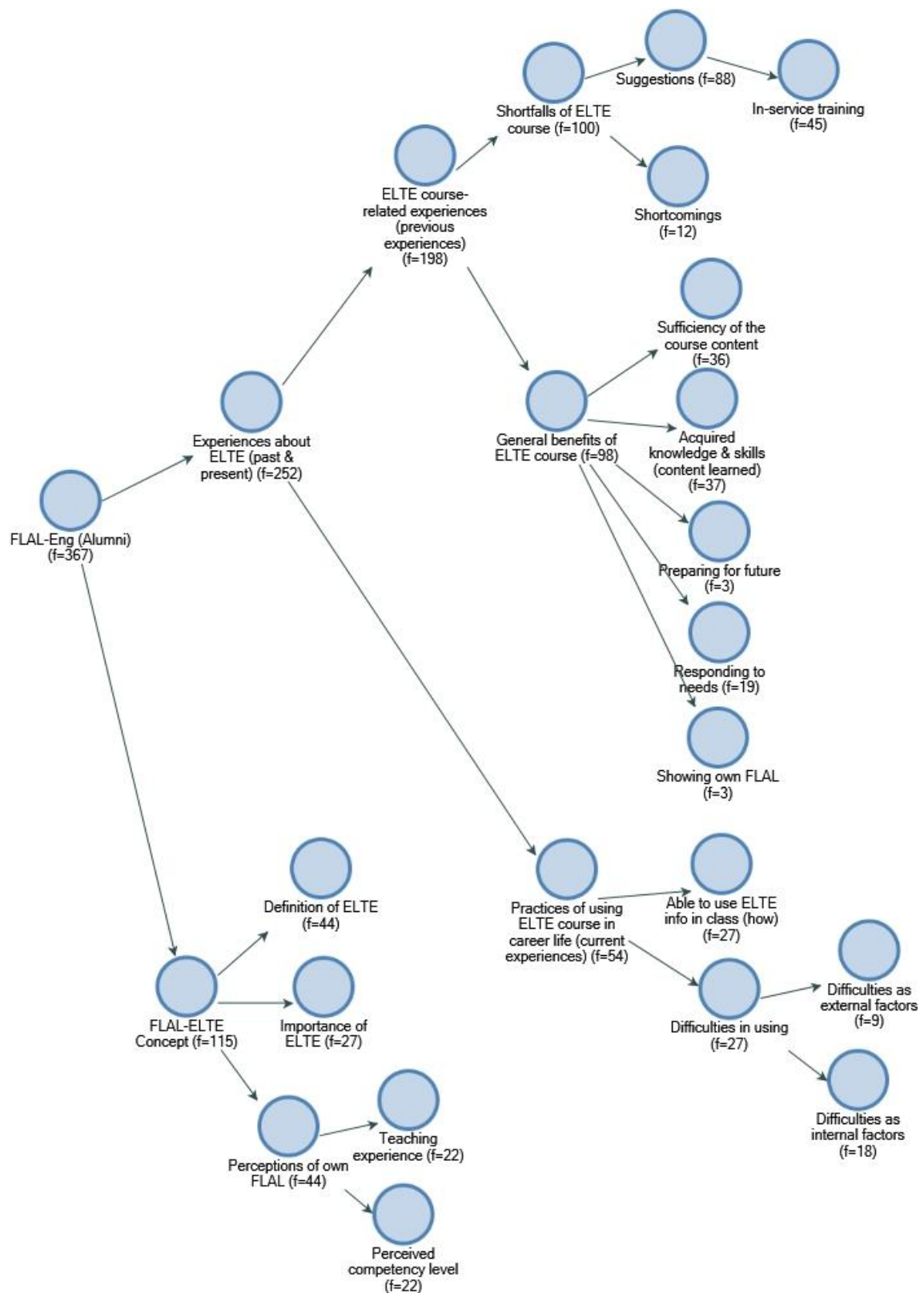


Figure 4.5. The thematic display of the findings from the novice EFL teachers

The comments indicated that novice EFL teachers mostly focused on certain keywords while discussing their opinions and experiences:

‘Experience, knowledge and skills, language skills and areas, language proficiency, measuring, testing, theoretical knowledge, practical side, ELTE course content, practicum coordination, in-service training, local conditions’

For *FLAL-ELTE Concept*, they gave definitions of ELTE ($f=44$), mentioned the importance of ELTE itself ($f=27$), and reported their perceptions of their own FLAL ($f=44$) (see Figure 4.6).

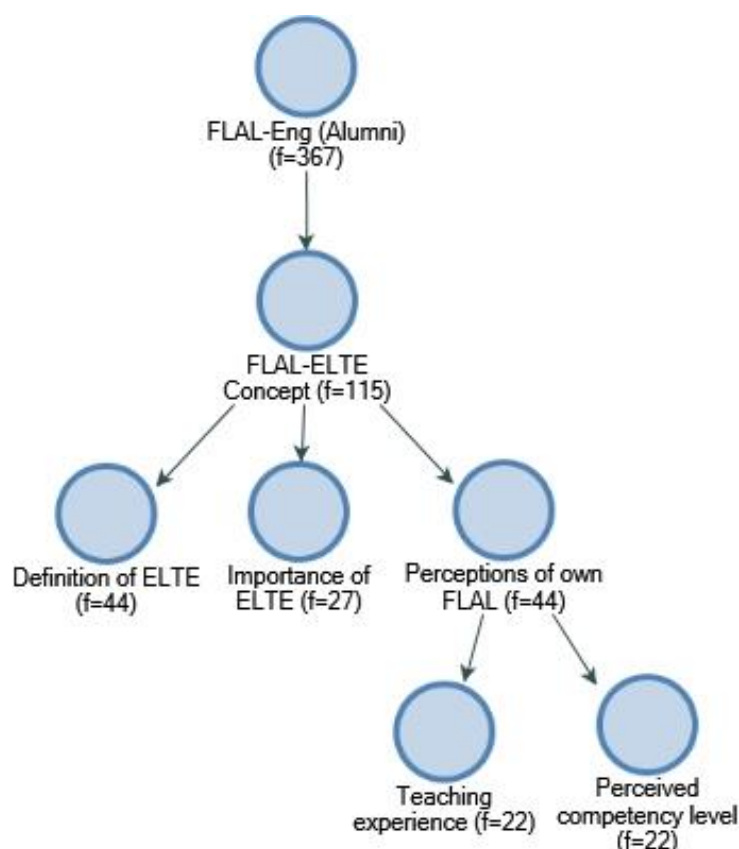


Figure 4.6. The thematic display of the theme ‘FLAL-ELTE Concept’ for novice EFL teachers

In the first place, novice EFL teachers defined what ELTE means. They mostly concentrated on the issue of language proficiency (language skills and areas) and the aims of language learning. Most of them indicated that language assessment means measuring and evaluating students’ progress and determining their deficiencies in the target language. For example, one noted that:

‘Assessment is to measure the proficiency level of language learners as well as evaluating their successes and progress in that foreign language.’ (Low.T.10)

Moreover, some of them focused on whether students are able to use language to express themselves rather than just considering their results in exams or tests. That is, according to them, ELTE refers to assessing students' progress in that language. For instance, one of them explained that:

‘Testing and evaluation in a foreign language is not only bound to tests or written exams, but also it means to assess students' ability to use the target language effectively.’ (High.T.8)

Besides, few novice EFL teachers defined ELTE from the perspective of planning and evaluating the whole language teaching/learning process in which they emphasized the objectives of language learning. One of them expressed that:

‘Testing-evaluation is the process of planning learning outcomes related to language skills, and to measure and evaluate to what extent these outcomes are achieved by means of testing tools.’ (High.T.11)

On the other hand, only one novice EFL teacher, who felt highly competent in FLAL, mentioned that teachers should have the knowledge and skills of ELTE itself; i.e. knowing how to construct exams, to interpret results and reporting them. S/he underlined that:

‘Language assessment means to have knowledge about the concepts and methods related to the testing of language skills, to examine test preparation processes related to language proficiency and development, to prepare a language test for a specific group of students, and to plan, implement and evaluate the results via item analysis. And also, it means to report assessment results in written format and present them verbally.’ (High.T.17)

Apart from these, novice EFL teachers focused on the importance of testing and evaluation in English language classes. Most of them underscored the fact that ELTE is important because of its significant role in testing language proficiency, determining and evaluating language progress, showing whether the aims are achieved, and getting feedback about deficient points in that language. For instance, one of novice EFL teachers pointed out that:

‘In the teaching/learning context, feedback should be given in order to evaluate and improve language proficiency of learners. So, feedback comes from the assessment results of learners, and is used again to develop their language proficiency. In this way, the shortfalls of the education may be fulfilled.’ (High.T.6)

In addition, some teachers argued that language assessment is an important procedure to assess whether their instruction is effective and to get feedback about the process itself. Therefore, one of them stated that:

‘Assessment plays an important role in terms of carrying on the instruction with new objectives if planned learning outcomes are achieved or going back to the study and making revisions if the objectives are not achieved.’ (Moderate.T.18)

However, only a few of novice EFL teachers drew attention to the multi-dimensional characteristics of foreign language learning; that is, language assessment should be tailored according to different language skills and components by using various techniques to assess. For example, one discussed that:

‘Assessment tools should be multi-faceted in ELT because language teaching cannot be thought as being only one dimension. The assessment procedures should be multi-dimensional including a student’s language progress with respect to his/her knowledge of grammar and vocabulary as well as the skills of listening and reading comprehension and the like.’ (Moderate.T.8)

As far as their perceptions of their own levels in FLAL; that is, how much qualified they felt in English language assessment, half of the participants perceived themselves to be highly competent whereas only two teachers felt themselves inadequate in terms of FLAL. When it comes to their teaching experience, most of them were very novice with two years of teaching English at different classes in Turkish state schools ($f=7$).

The other theme revealed from the findings of novice EFL teachers’ statements is ‘Experiences about ELTE (past & present)’. For *Experiences about ELTE (past & present)*, they highlighted their previous experiences about ELTE course ($f=198$), and their current experiences about using their knowledge and skills in their professional lives ($f=47$) under two categories.

The first subcategory is about their ‘*previous experiences about ELTE course*’ they took in their ELT training program. On the one hand, they discussed general benefits of the course for their teaching life ($f=98$), and on the other hand, they complained about the shortcomings of the ELTE course while making some recommendations to improve it ($f=100$) (see Figure 4.7).

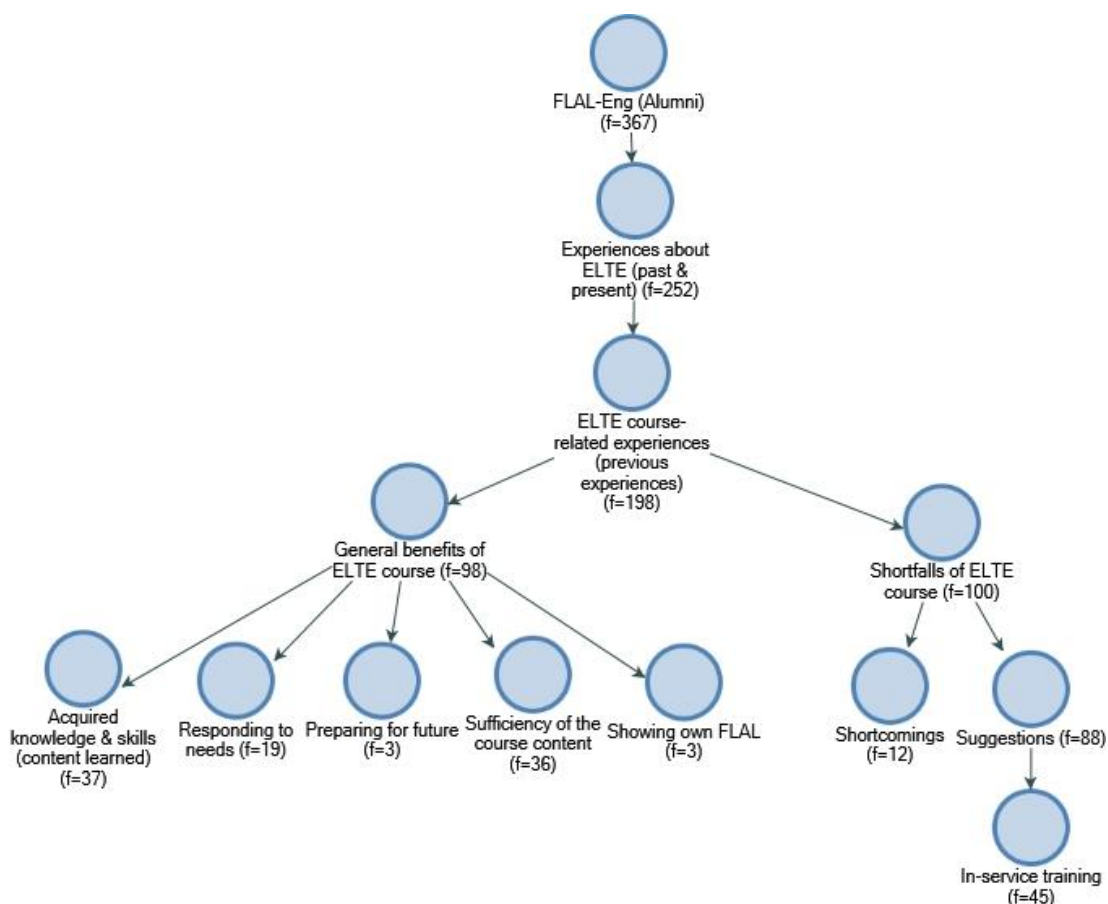


Figure 4.7. *The thematic display of the subtheme ‘ELTE Course-related Experiences (Previous Experiences)’ for novice EFL teachers*

To begin with general benefits of ELTE course, they mentioned their acquired knowledge and skills about language assessment ($f=37$), whether the content is sufficient ($f=36$), the feature of the course that it prepares undergraduates for the future teaching life ($f=3$), whether the course met their needs or not ($f=19$), and lastly, their own FLAL as a sign of professional competence ($f=3$). With regard to what they learned during ELTE course when they were studying ELTE at university, they highlighted the ability to assess language skills and components in a planned way by using several tools, and they also noted the criteria to be considered for effective testing procedures (e.g., reliability, validity, age, levels, etc.). For instance, one of them exemplified that:

“Surely, this course [ELTE in teacher training] has helped me a lot. Simply put, it has been helpful with respect to such topics as ‘how a language test should be designed, what it should be covered, how the questions/tasks should be organized, and how the interpretation/evaluation should be made’”. (Moderate.T.7)

Furthermore, one of the novice EFL teachers, who felt highly competent, focused on his/her ability to prepare language tests according to students' levels and course aims on the contrary to his/her colleagues who used available exams regardless of such points. S/he reported that:

‘Considering my first experience [about language assessment], I can say that I have not had any difficulty in testing language [at my school]. I am able to construct my own assessment criteria. Although the other teachers in the school [colleagues] prefer ready-made assessment tools, I created my own tests. Since my tests are more appropriate to my students as well as the learning outcomes, I have achieved more success.’ (High.T.4)

Pertaining to the sufficiency of ELTE course, all of the participant teachers agreed that the course content was adequate and suitable for their professional life. For instance, one emphasized that:

‘The content of the course [ELTE course] consisted of real-life assessment implementations and the undergraduates were very active in that course. Therefore, I think the course is very beneficial and advantageous. In general, the course was sufficient.’ (High.T.4)

Accordingly, they pointed out that ELTE course was beneficial because it was very useful to prepare them for their career. For example, one of them stated that:

‘The level of sufficiency [of ELTE course] is appropriate to the students who will be teachers after graduation.’ (Moderate.T.5)

In terms of novice EFL teachers' expectations and needs, nearly all of them were satisfied with ELTE course whereas some of them indicated some lacking points of the course such as little practice. They discussed that:

‘I think the course [ELTE course] was adequate. It provided us with necessary knowledge and skills about assessing language skills.’ (Moderate.T.2)

‘I cannot say that this course [ELTE course] did meet my needs and expectations. I only gained the information about assessment evaluation, but I think the course was deficient in terms of practice.’ (Moderate.T.20)

On other hand, even though most of the participant teachers agreed that ELTE course was helpful and sufficient, they criticized the course being very theoretical and it should have included more practical exercises and implementations. Few of them also underlined that the class hours were not adequate, and the last semester was too late to take the course. For instance, one commented on the shortcomings like that:

‘The content of the course [ELTE course] was satisfying in terms of theory but the practical side was not enough.’ (Moderate.T.20)

Therefore, they made a number of suggestions how to improve course itself. Most of them highlighted the deficient side of the course which was practice. Regardless of their perceived levels in FLAL, teachers said they needed more practice to enhance their testing and assessment skills. Thus, they suggested more practical tasks on assessing language should be added, and students should be made to prepare language test items on their own. They also mentioned that new topics such as analyzing test results can be added, and some simulations for performing tests can be included. For example, one recommended that:

‘The [ELTE] course was very theoretical. More practice would have been better. For example, for each language skill, students should prepare an assessment tool every week, and the criteria for the evaluation should be interpreted.’ (High.T.1)

Moreover, some of novice EFL teachers complained that they did not learn anything about the evaluation side of testing such as scoring and feedback. Thus, one of them suggested that:

‘[To the ELTE course content], exercises of preparing various test items, techniques of evaluation and analysis, and statistical calculations [can be added as new topics]’. (Moderate.T.2)

Besides, some novice EFL teachers approached the practical issue from the point of practicum. According to them, undergraduate students should be made to prepare and administer some tests in their teaching practice for the sake of gaining experience in assessment in addition to teaching. Therefore, one of them suggested that:

‘Before teacher trainees start teaching profession, they should be provided with experience [practice] in terms of assessment process by making them to prepare at least a part of a language exam during micro-macro training [practicum courses].’ (Moderate.T.5)

Similarly, one of the teachers, who felt less competent in FLAL, drew attention to certain drawbacks of ELTE course such as class hours, semester time and practical side. Therefore, s/he made some recommendations like:

‘Various different testing and assessment methods for listening, writing, reading, speaking and grammar as well as the practices of them should be given in the [ELTE] course. In addition, students should prepare test items accordingly and should have an opportunity to apply those items and evaluate them. The hours allocated to the course should be increased. The course should be given for two semesters: In the first semester, the theoretical side of assessment should be taught, and in the second semester, the practical side should be applied. Also, the practice of assessment for each skill can be integrated into the practicum in the

second semester. Hence, students will be able to practice assessment in the real classrooms.’
(Low.T.10)

On the other hand, the same teacher criticized ELTE course in terms of the content about assessing productive skills, which was very sufficient and deserved to be emphasized more than grammar and other receptive skills. It is because the nature of assessing productive language skills includes subjectivity. Hence, s/he discussed that:

‘I think I acquired the essential knowledge about assessing grammar, listening and reading. But I think the [ELTE] course was not successful enough in terms of the topics about assessing writing and speaking. [...] Especially, I cannot assess students’ speaking and writing skills. In my opinion, the knowledge about assessing speaking and writing skills that I gained at university is inadequate. It is because there was not much practice about that topic.’ (Low.T.10)

Among the suggestions made to improve ELTE better, in-service training is very notable because they believed that some of the missing parts of the course can be covered, and their own inabilities about language assessment can be eliminated with the help of in-service training. For instance, they assumed that in-service training might be very helpful to remind them the previous ELTE knowledge and practices they learned because they complained they forgot certain things about assessment throughout years. One of them stated that:

‘I think in-service training [about ELTE] is necessary because in-service teachers prepare language exams without paying attention to some assessment criteria after a while. Therefore, constantly, in-service seminars should be provided suitable to the changing needs of the system and what is learned at university should be kept fresh all the time.’ (Moderate.T.12)

Likewise, some of the teachers thought that professional teaching competence also requires to be competent in language assessment; and thus, they recommended that several different testing and assessment techniques can be shared and discussed for different ages and levels of language learners in in-service training sessions. One pointed out that:

‘There might be in-service training. Specifically, it may be about assessing different language skills in different language school contexts as well as the difficulties encountered while assessing.’ (Low.T.13)

Other than these recommendations about in-service training, some novice EFL teachers focused on the education of ELTE, and they put forward that not all teachers had such kind of education at their universities and they were very incompetent in terms of

language assessment. They also criticized that even those who took ELTE training, not all universities took care of the course at the expected level and extent. Therefore, they agreed that in-service training was very important to help those teachers about language assessment. They explained that:

‘In-service training [about ELTE] should be provided. I think each university does not put emphasis on this topic in the same way. I have come to this point on the basis of the materials that are prepared by my colleagues.’ (High.T.19)

‘In-service training [about ELTE] is absolutely necessary. There are the colleagues who did not have any training about language assessment at my school. Teachers, especially those who make practical exams in their lessons, should be equipped with such knowledge and skills.’ (Moderate.T.6)

The second subtheme of ‘Experiences about ELTE (past & present)’ is about novice EFL teachers’ practices of using their learned assessment knowledge and skills in their teaching practices as their current experiences ($f=54$). While some of them believed that they were able to assess properly their students in the light of ELTE knowledge ($f=27$), some other underlined their difficulties in practicing assessment due to either external ($f=9$) or internal factors ($f=18$) (see Figure 4.8).

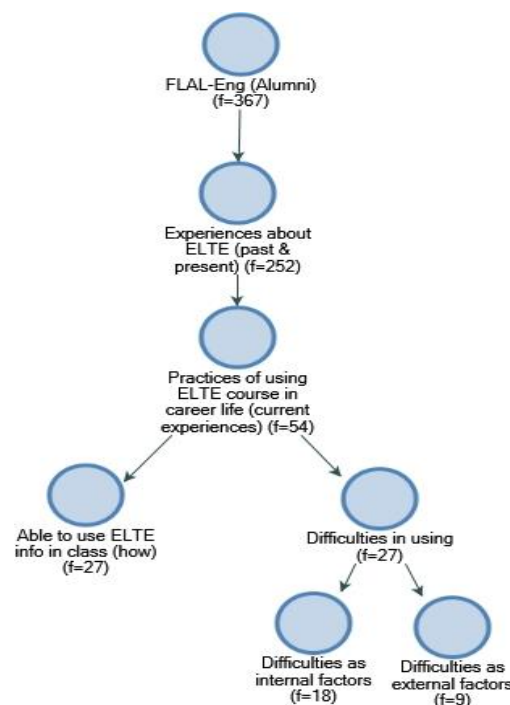


Figure 4.8. *The thematic display of the subtheme ‘Practices of Using ELTE in Career Life (Current Experiences)’ for novice EFL teachers*

Regarding novice EFL teachers' ability to use their knowledge and practices they learned in the ELTE course, most of them gave specific examples how they used such skills to assess their students' language proficiency. They mostly concentrated on the test development criteria and techniques to measure students' proficiency rather than how to monitor students' progress in the target language during the teaching process. One of them exemplified such that:

'It did work [ELTE course was helpful at work]. For example, I pay attention to the difficulty level of test items while preparing exams. I avoid giving clues about correct answer in true-false test items. I try to be careful about the clarity of questions. I prepared each test item which measures only one skill. I pay attention to the length of the gaps in the fill-in-the-blanks test items.' (High.T.8)

When it comes to the difficulties the participant teachers had in language assessment, they indicated that external factors as well as internal factors affected their assessment procedures and performances. To begin with external factors, they underlined the issue of local contexts where they work as teachers; for example, they stated that their students had lower levels of English language proficiency even though their age and education levels were high; and thus, they could not apply high-level tests to their students. For instance, one pointed out that:

'I have some difficulties in assessing English. The proficiency level is very low in my high school where I work, and thus, I cannot use ready-made tools. I write all my test items on my own because students have problems of critical proficiency levels.' (Moderate.T.22)

Moreover, one of the novice EFL teachers drew attention to Turkish examination system which is based on high-stakes testing in the form of multiple-choice questions; and hence, such exams have the priority in Turkey. Therefore, s/he must conduct written exams like that in his/her classrooms. S/he expressed that:

'In fact, this [ELTE] course is very good for teaching profession. All in all, written examinations have an important place in our country [Turkey].' (Moderate.T.3)

In the same vein, one of the other novice EFL teachers mentioned Turkish status about foreign language context in which English is learned as a foreign language. S/he emphasized that Turkey has a foreign language learning context; and therefore, students have no chance to use the target language in their daily lives much. Hence, it is very difficult to assess the use of language outside of the class and limited in terms of testing language skills in the classes. S/he explained that:

‘If we talk about Turkey, it is crucial to assess to what extent we can teach English in the classroom to those who are mother-tongue is not English, and also do not have any chance to hear and use English in their daily lives.’ (High.T.19)

As for the internal factors that challenge novice EFL teachers in language assessment, most of them focused on the lack of practice and experience in assessing students’ language progress. They complained about their inability to perform effective testing practices in their current classes because they were unable to adapt their theoretical knowledge to their teaching contexts. For example, one argued that:

‘I have difficulties in practice and implementation. I have theoretical knowledge [about language assessment], but I have difficulty in combining such knowledge with the teaching style and testing system at state schools [in Turkey].’ (Moderate.T.20)

To sum up, it can be concluded that most of novice EFL teachers had apparently great knowledge of ELTE thanks to the course, but they had some difficulties in practicing due to lack of experience. While they successfully defined the concept of ELTE and believed the importance of assessment, they mostly emphasized the notion of summative assessment both in terms of the definitions and their current uses of assessment in teaching contexts.

4.1.3. Findings from pre-service ELT teachers

Pre-service ELT teachers taking ELTE course at Anadolu University took part in the study to express their opinions and share their experiences about the course as well as their general conceptions about language assessment. They responded to open-ended questions two times: Before taking the course and after the course.

- Before the course -

At the beginning of the semester, 164 course-takers participated in the research. The analysis of their responses before taking ELTE course yielded two themes: *FLAL-ELTE Concept*, and *ELTE Course-related Issues* (see Figure 4.9). Pre-service ELT teachers commented more on ELTE course-related issues ($f=1046$) than FLAL-ELTE concept ($f=253$) (see Appendix-17 for detailed tabulation).

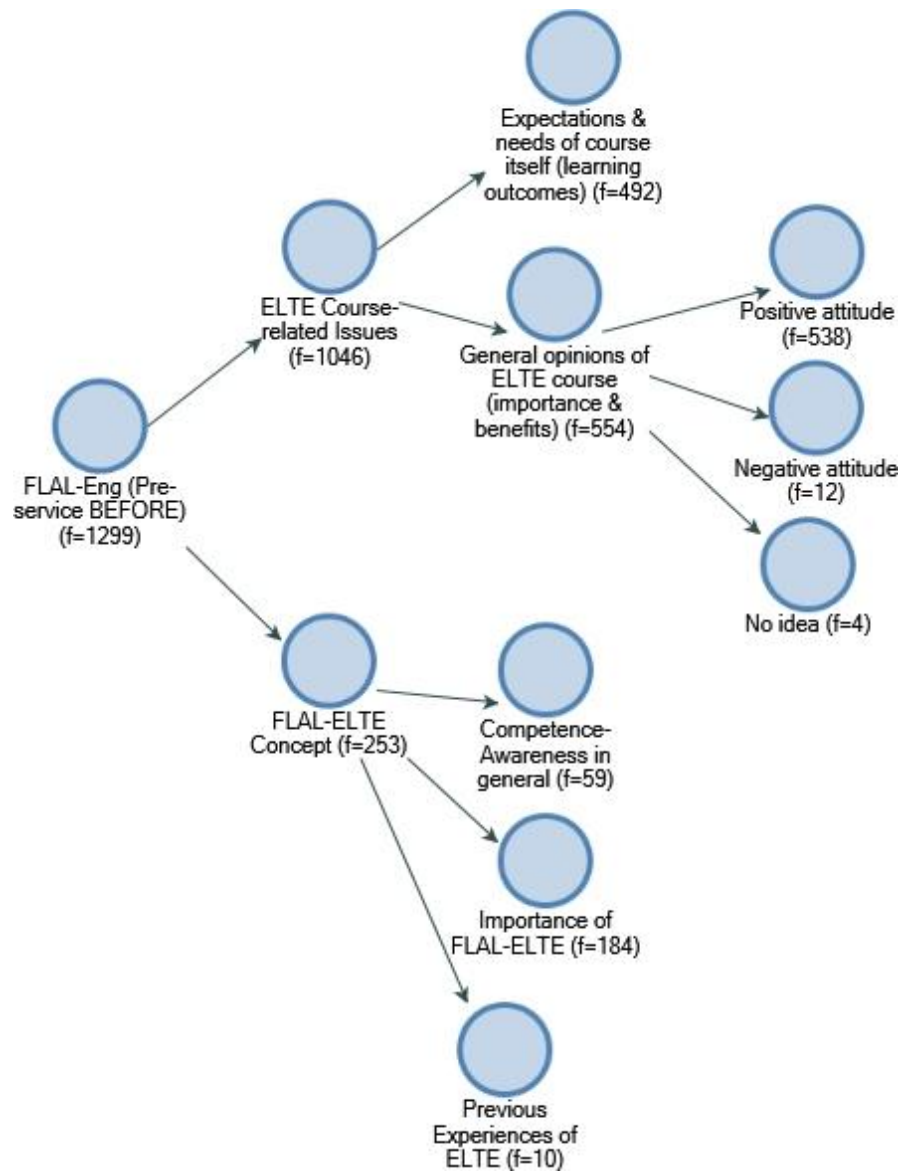


Figure 4.9. The thematic display of the findings from pre-service ELT teachers (before the course)

The comments indicated that pre-service ELT teachers before they took the course mostly focused on certain keywords while discussing their opinions and experiences:

‘ELTE course content, expectations and needs, importance and benefits, teaching career, properly testing, exam, learn, knowledge and ability, competent’

For *FLAL-ELTE Concept*, they underlined what kind of competencies are essential for English language assessment ($f=59$), mentioned the importance of FLAL-ELTE itself ($f=184$), and shared their opinions about the previous experiences of own assessment related issues ($f=10$) (see Figure 4.10).

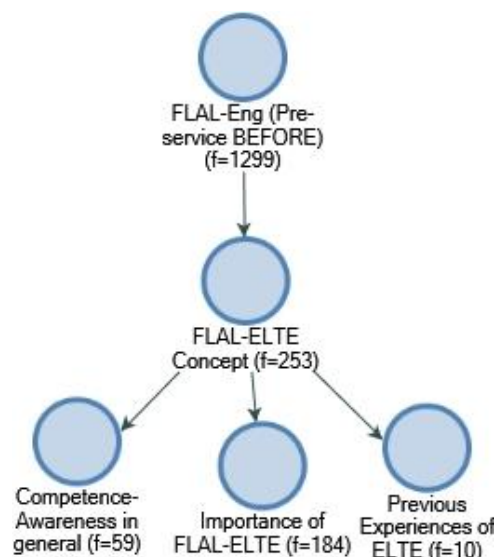


Figure 4.10. *The thematic display of the theme ‘FLAL-ELTE Concept’ for pre-service ELT teachers (before the course)*

Firstly, most of the pre-service ELT teachers underlined the fact that language teachers need to be competent in how to test and assess students’ knowledge of the target language. They pointed out that language teachers must know not only how to teach but also how to test since teaching and assessment are embedded in each other. Therefore, they believed that teacher candidates should gain necessary competencies and skills about language assessment during their teacher education. For example, one commented like:

‘Assessment in ELT is necessary. Being a teacher is not just about teaching. One of the most important points that every teacher has to pay attention is to prepare language exams suitable for their students. Therefore, every teacher candidate needs to be competent in language assessment.’ (Before.Pre.23)

Secondly, all of the pre-service teachers underscored the importance of language testing and assessment in teaching life. They stated that assessment is as important as teaching itself, and it is mostly helpful to find out whether students learn or not and to get feedback about the effectiveness of teaching process. One of them highlighted that:

‘I think this course is necessary and should be given. I believe that every teacher needs to know how to teach, how to evaluate the quality of their teaching, how to assess to what extent their teaching activities are successful, and how to test how much of the goals of their teaching activities are achieved. Teachers should be well-equipped with that kind of knowledge and skills.’ (Before.Pre.77)

Lastly, some of the pre-service ELT teachers mentioned their previous experiences about testing and assessment as test-takers. Mostly, they complained about taking poorly prepared and wrong exams, which had great impacts on their educational decisions, throughout their school lives. They also reported that their previous English language teachers did not know how to test properly. Thus, one of them criticized that:

‘My expectation from this course is to study on good test samples rather than poor test items. We have already come across poor tests, and still we have seen them. I think it is necessary to study on good test items to construct a good perception of assessment.’ (Before.Pre.49)

The other theme emerged from the data of pre-service teachers before taking the course is *ELTE Course-related Issues*. In this category, they focused on their expectations and needs about the course itself ($f=492$) and expressed their general opinions about the course ($f=554$) (see Figure 4.11).

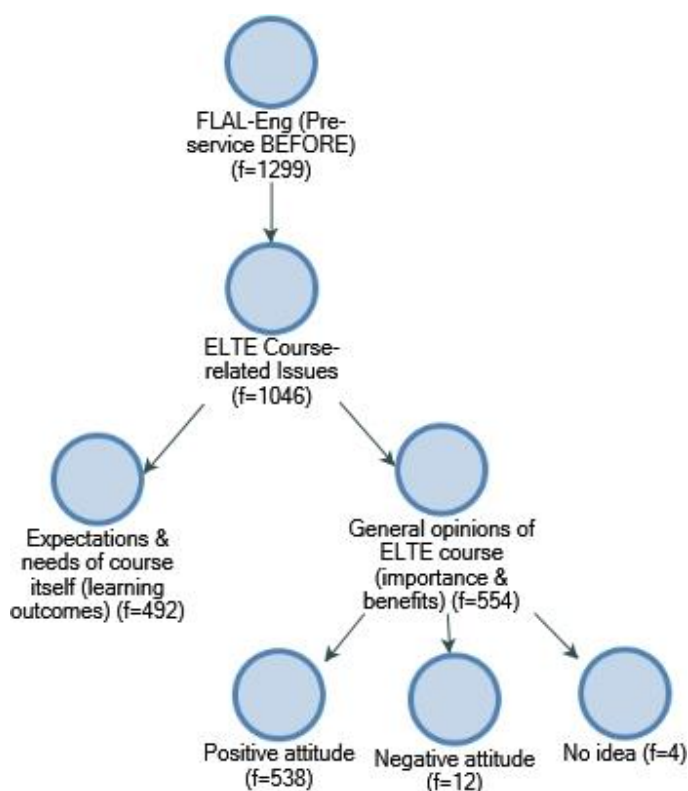


Figure 4.11. The thematic display of the theme ‘ELTE Course-related Issues’ for pre-service ELT teachers (before the course)

With reference to the expectations and needs of pre-service teachers about ELTE course, most of them hoped to learn how to assess language and be able to effectively test and evaluate language skills during the course period. They also indicated that they

expected to gain knowledge and ability of assessment such as preparation, designing, and evaluating language tests according to certain criteria. In short, they believed they would be competent in terms of all the concepts of English related testing and evaluation. For example, one of them stated that:

‘I believe this course will help me to obtain adequate knowledge and skills about preparing, administering, and evaluating tests.’ (Before.Pre.19)

Furthermore, some of the pre-service teachers highlighted that they needed to be good at language testing because they would be teachers and wanted to use assessment methods effectively and properly to provide better language education in the future. One pointed out that:

‘In this course, I think we will learn how to assess language learners. I expect the course will be relatively permanent, and I want to gain the knowledge that I will use in my professional teaching life.’ (Before.Pre.28)

In addition, some teacher candidates highlighted that this course would be a revision of English teaching methods-courses they took for three years and they hoped to relate their methodological knowledge with assessment knowledge for their future professional life. For instance, one noted that:

‘[At the end of this course], I think I will be able to teach and assess properly by combining teaching methodology knowledge as well as professional knowledge that I learned [throughout training years]. (Before.Pre.57)

Besides, some of the course-takers expected to gain more awareness about the issue of testing and assessment by referring to their previous testing experiences, which would make them to criticize the features and qualities of language tests. For this thought, one stated that:

‘Throughout 4-year teacher training, have all the exams we have taken so far been implemented correctly or not? I think I will learn some kind of knowledge to help to evaluate such an issue.’ (Before.Pre.150)

However, only a few of prospective teachers drew attention to the concept of backwash effect of testing, and they discussed what sort of affective and psychological effects the language test may have on students as well as themselves. To exemplify, one argued that:

‘I think I will be a knowledgeable teacher about language assessment. In addition, I will be able to make a connection between what I teach and assess as well as to see the psychological effects of tests on students [backwash effect].’ (Before.Pre.50)

Unlike these findings and statements, a few of pre-service teachers expected only to enjoy and pass the course because they were last-year students and would be graduated soon. They stated that:

‘Since I am a senior student, my first expectation is to pass the course and to learn something [about assessment]. (Before.Pre.111)

As for the general opinions of pre-service teachers about whether ELTE course is beneficial and important, nearly all of them had positive attitudes towards the course. They emphasized the importance and benefit of the course for their professional life because they believed they would encounter problems while assessing their students and their teaching if they did not learn and gain necessary knowledge of ELTE. It is also because they attached more importance to classroom and formative assessment due to their need to test perpetually their future students’ language progress on their own. Therefore, one of them reported that:

‘I think this course should be given. It is necessary for pre-service teachers studying at faculty of education in order to learn and know how to assess language skills and the evaluation criteria for a successful and qualified professional life.’ (Before.Pre.7)

Moreover, some of the pre-service teachers realized that only teaching is not enough; teachers must have the competence of assessment to make further developments in their teaching by improving themselves. They also underscored the fact that from first year of the ELT program to the last year, they learned how to teach English, but it is time to learn how teach in order to help their students when they begin to teach. One indicated that:

‘I think this course should be taught because so far, we have learned how to teach. At this last year, we will learn how to test. What did we learn, which level are we at, and have we come to the level of making evaluations? This course may answer such questions.’ (Before.Pre.35)

Similarly, some of them believed that this course would be useful to increase awareness and lead to the positive change of perspectives about language assessment. One pointed out that:

‘I believe this course will be beneficial. It will change our point of view towards testing and we will understand how to give a better education by breaking down our prejudices.’ (Before.Pre.165)

In addition to these, some of the pre-service teachers focused on the Turkish examination system and they thought that exams are part of their lives. They criticized that the exams are very problematic while they have a great impact on test-takers. Therefore, this course is hoped to be helpful to increase their knowledge and skills of assessing language proficiency as well as to gain awareness about testing. One discussed that:

‘In my opinion, testing and evaluation are one of the most important stages of educational process. Assessment is a feedback both for teachers and students. While students have the chance to see their mistakes and learn the correct ones, teachers evaluate their teaching techniques [as a result of assessment]. Furthermore, since every stage of educational system in Turkey is dependent on exams, every teacher should know how to prepare tests. Therefore, this course is important and necessary.’ (Before.Pre.58)

In another respect, some undergraduate students compared Turkish general testing and evaluation course with English-oriented one. They argued that the general course, as the name suggests, is suitable to every area of education without the consideration of testing language skills. Thus, they believed ELTE course is very important because it focuses specifically on English-related assessment and evaluation. In this sense, one stated that:

‘This course is important because the course of assessment and evaluation instructed in Turkish [at the third year of training as general educational and pedagogical course] is very general. But we teach a completely different language [English as a foreign language] and apart from general practices, there must be different ways to assess English language.’ (Before.Pre.104)

On the contrary to positive attitudes towards ELTE course, a few of the participant teacher candidates held negative attitudes especially because they questioned the quality of this training in terms of whether pre-service teachers are well-equipped with required knowledge of testing. For instance, one of them doubted that:

‘Sure, it is a necessary course for every teacher. But when we consider the conditions of Turkey, some questions arise: Has each pre-service teacher taken this course?; if they have taken, how effective has the course been?; after graduation, have they fulfilled the requirements of this course in their teaching? I think it is an awful condition.’ (Before.Pre.63)

All in all, pre-service teachers before taking ELTE course expected to acquire necessary knowledge and skills of assessment for their future career; and hence, they

believed this course would be beneficial because assessment is integral part of teaching and because the notion of testing is very important in Turkey.

- After the course -

As far as the end of the course semester is considered, 146 course-takers responded to the questions. The analysis of their responses after ELTE course produced two themes: *FLAL-ELTE Concept*, and *ELTE Course-related Issues* (see Figure 4.12). Pre-service ELT teachers shared their perceptions of ELTE course-related issues ($f=1036$) more than FLAL-ELTE concept ($f=617$) (see Appendix-18 for detailed tabulation).

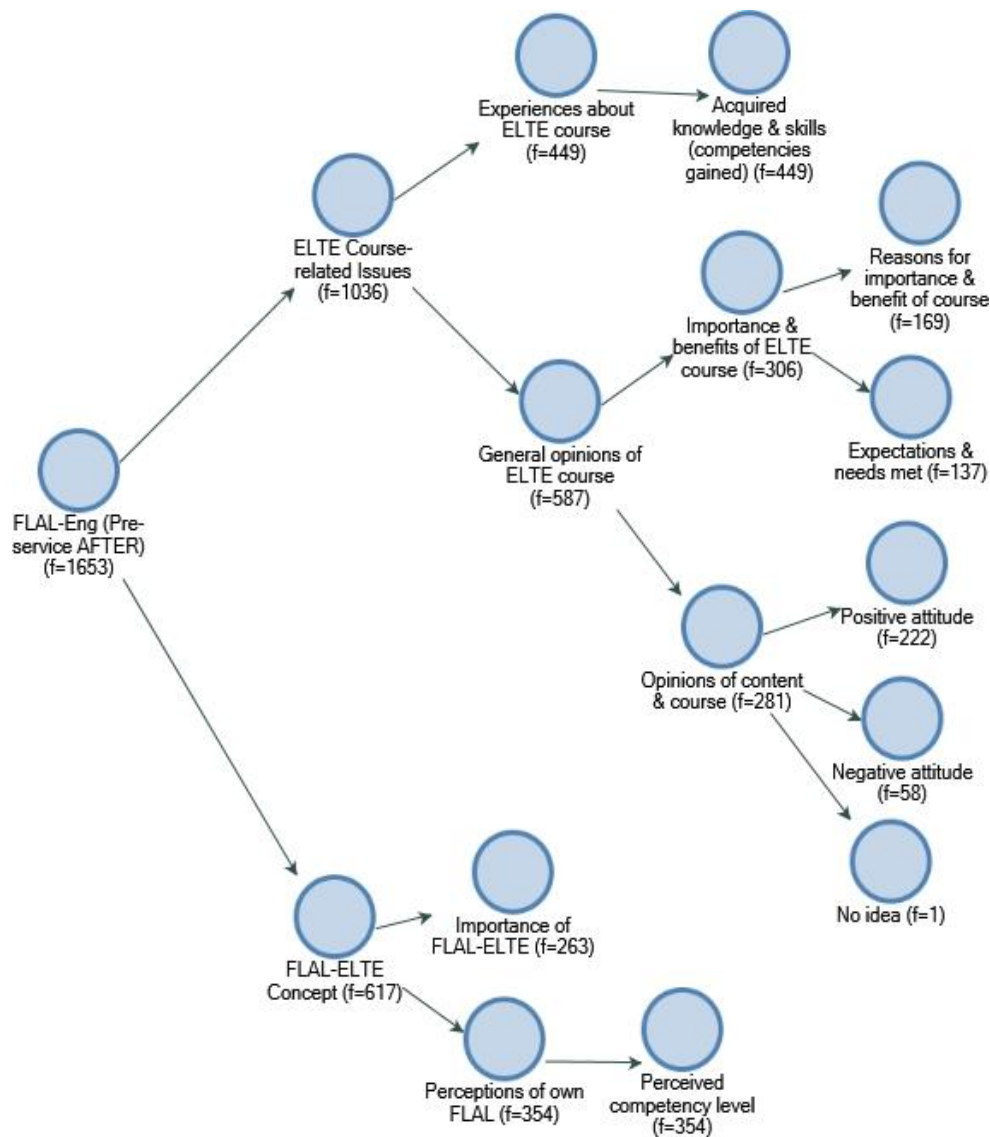


Figure 4.12. The thematic display of the findings from pre-service ELT teachers (after the course)

The comments indicated that pre-service ELT teachers after they took the course mostly focused on certain keywords while discussing their opinions and experiences:

‘ELTE course, competencies gained, awareness, criteria and principles, types and items, theory much, practice less, knowledge and skills, specific competence, language proficiency testing, moderate, future career, update, topics, practicum’

For *FLAL-ELTE Concept*, they underscored the importance of FLAL-ELTE in general ($f=263$) and mentioned their perceptions of their own competency level in ELTE ($f=354$) (see Figure 4.13).

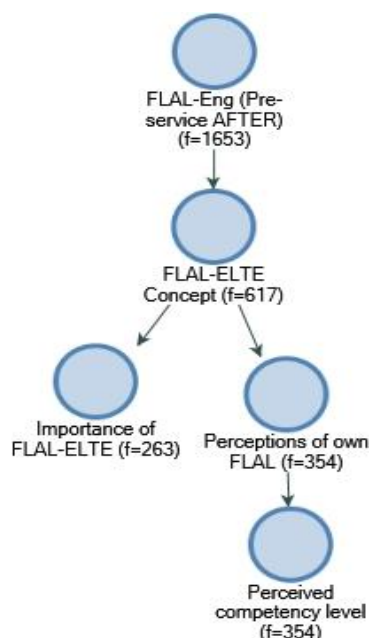


Figure 4.13. *The thematic display of the theme ‘FLAL-ELTE Concept’ for pre-service ELT teachers (after the course)*

Considering the importance of language assessment in general, most of the pre-service teachers thought that assessment is as important as teaching itself because testing and evaluation are helpful to detect whether students learn or not. It is also useful to test whether the instruction is effective by getting feedback about the process. Therefore, they stated that every teacher, no matter what they are candidates, novice or experienced, must acquire this specific competence. To highlight these aspects, one pointed out that:

‘One of the most important competencies that every teacher must possess is preparing language exams. Teaching is not the only important thing, but at the same time, testing/assessment is very important. I believe we will be able to assess properly thanks to this course.’ (After.High.Pre.7)

In addition, few of them emphasized that ELTE is complementary to what they learned during the ELT program. That is, they expressed that knowing only to teach is not sufficient but also knowing how to test is important since assessment is integrated into teaching. Thus, one reported that:

‘In fact, this course is complementary to what we have learned so far [throughout our teacher training]. It is because it is very important to test what we teach, to assess the progress of students’ language development, and to supply learning/teaching process.’
(After.Moderate.Pre.113)

As for the perceptions of their own FLAL levels, most of pre-service teachers believed that they needed to improve themselves more in terms of language assessment seeing that they felt themselves competent at the moderate level in language assessment after the course ($n=109$). They hoped to make progress in testing with respect to especially practical skills when gaining more experience in teaching in their future life. Hence, one of them commented like that:

‘I perceived myself being qualified in language assessment at a moderate level. It is because I cannot foresee how effectively I will put theoretical knowledge into practice. I believe I will be better in language assessment via experience.’ (After.Moderate.Pre.100)

By contrast with the moderate level pre-service teachers, those who perceived themselves highly competent in FLAL ($n=33$) discussed that though they were still teacher candidates, they assumed to learn necessary things about language testing and evaluation and be able to perform tests easily in their teaching life thanks to this course. In this view, one of them pinpointed that:

‘I perceive myself more competent now [at the end of the semester]. I noticed my mistakes while preparing language test items or activities, and I learned how to fix them.’ (After.High.Pre.90)

However, though there were few teacher candidates who perceived themselves to be very inadequate ($n=4$), they complained about their lack of practice and experience about English language testing and evaluation even after taking that course. Therefore, one indicated that:

‘I do not perceive myself being competent in language assessment because of the lack of experience. But I hope this situation will change after starting my career as an English teacher. I believe I will be better by means of practicing.’ (After.Low.Pre.78)

The other theme is about pre-service teachers’ general thoughts about the ELTE course they took. For *ELTE Course-related Issues*, they expressed their opinions about

the course (importance and content) ($f=587$) and mentioned their experiences throughout the course in the form of what they learned ($f=449$) (see Figure 4.14).

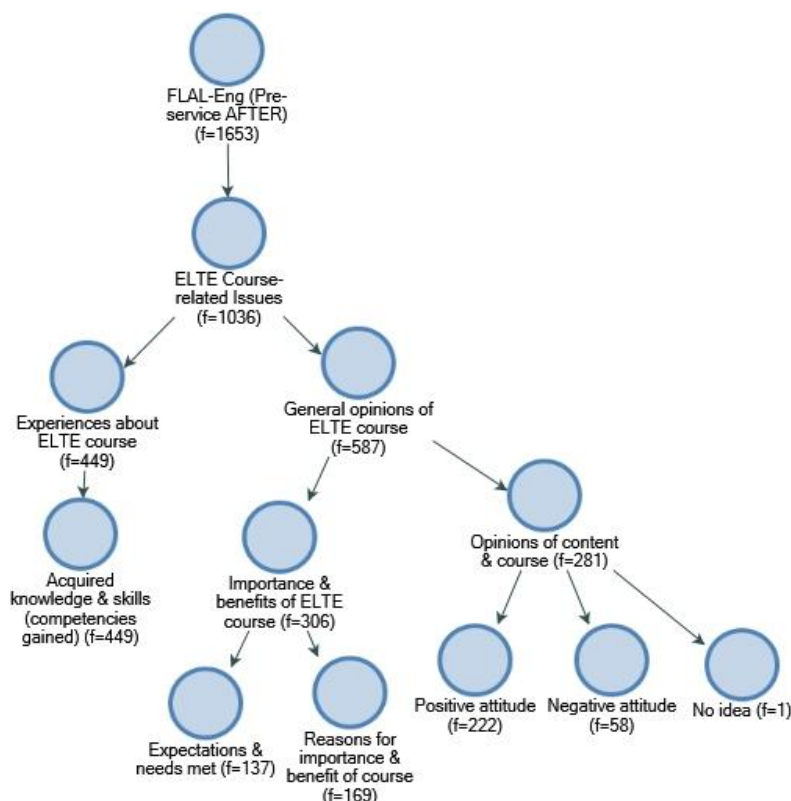


Figure 4.14. *The thematic display of the theme ‘ELTE Course-related Issues’ for pre-service ELT teachers (after the course)*

Concerning the experiences about ELTE course, pre-service teachers focused on their gained competencies about language assessment in which they discussed what they learned and how they could use it. They reported that they learned how to test each language skill as well as how to design such tests by using different test items appropriately. For instance, one of them pointed out that:

‘At the end of the course, I learned how to assess language skills (grammar, vocabulary, listening, etc.) and also, the criteria which I should bear in mind while assessing.’
(After.Moderate.Pre.100)

Besides, most of the pre-service teachers highlighted certain topics such as table of specifications, the criteria and principles of English language testing, types of tests and the like. One of them exemplified that:

‘We received information about the criteria that we have to pay attention while designing language tests. Also, we learned how we can prepare language exams according to the level

of our classes as well as our teaching goals. Besides, we studied on the procedures which can be utilized to increase the reliability and validity of language tests.’ (After.Moderate.Pre.119)

Other than these, some of the pre-service teachers stated that their views of testing and assessment changed; they said they realized teaching-testing connection and many affecting factors on testing. For instance, one indicated that:

‘I learned at the end of this course that there is a close relationship between teaching and assessment. I also learned about language testing techniques, evaluation, the features that any measurement tool should possess, and how to prepare test items for language skills such as grammar, vocabulary, reading, writing, and listening.’ (After.Moderate.Pre.44)

When it comes to general opinions about ELTE course, pre-service teachers mentioned the benefits and importance of the course, whether their expectations and needs were met, and what was good and what went wrong during the course. Nearly all of them agreed that the course responded to their needs and expectations, and it was beneficial. It is because this course was found to be useful for the career as future teachers and they believed they gained necessary competencies about language assessment. For instance, one stated that:

‘As a language teacher candidate, I learned about what sort of aspects that I have to consider while designing language tests, how to prepare and evaluate test items, and how to give feedback in my future teaching career. That is why, this course was beneficial.’ (After.Moderate.Pre.127)

Moreover, some of them indicated that ELTE course complemented to the knowledge of methods-courses that pre-service teachers took during the four-year training. Therefore, they thought that they acquired another competence, namely LAL, about professional teaching. Therefore, they became more aware of ELTE. For instance, one of them discussed that:

‘This course met my expectations. Throughout our 4-year-university life, we took the courses about how to teach English language; I mean, how to organize language teaching activities in the class. But we had not known how to assess those at the end of those courses. From this point, this course filled this gap.’ (After. Moderate.Pre.31)

However, few of them highlighted their ability to transfer knowledge between two courses about testing and assessment. That is, they argued that they were able to use their knowledge of general Turkish educational assessment course in order to supply ELTE knowledge. In this sense, one of them pointed out that:

‘This course was helpful. I took a similar version of this course in the third year, but it was Turkish and general [Measurement and Evaluation in Education]. Thus, by taking this specific course dedicated to English language testing and evaluation, I believe I will be more effective in terms of language assessment in my professional life.’ (After.Moderate.Pre.87)

As far as the opinions of the course itself and the content, nearly all of the pre-service teachers held positive attitudes in that they found course beneficial and the content sufficient. They emphasized that the topics were comprehensive, and they learned test types and other stuff about assessment while developing an understanding of the concept. For instance, one of them stated that:

‘I think the course content was sufficient. Providing detailed theoretical and practical knowledge about assessing language skills made us understand the concept of assessment, which was pretty good.’ (After.Moderate.Pre.83)

Besides, some of them underlined the topic of testing language skills. They concluded that they learned test types, sample test items and assessment methods for each language skill. For example, one illustrated that:

‘I think the course content was sufficient because there were adequate options for assessment examples for each language area and skill, and also, we learned what students would need as a result of assessment.’ (After.Moderate.Pre.112)

Another point highlighted by pre-service teachers about the course was that their course instructors were very qualified in testing and assessment; and hence, they were very helpful and explanative throughout the course semester. To exemplify, one noted that:

‘The course responded to my needs and expectations. I think I had learned a lot because of our course lecturer whose teaching was very good, clear, comprehensible and explanative.’ (After.Moderate.Pre.100)

On the contrary to positive comments, there were some pre-service teachers who discussed the shortcomings of the course and made certain suggestions to fix them. They mostly criticized the issue of theoretical side of the course because they stated that they had no practice in terms of testing language skills during the term. They stressed that giving only final assignment, preparing a language exam, did not guarantee the practice. Thus, one argued that:

‘The course did not respond to my needs adequately because the whole semester was theoretical. The practical part was given only as an assignment. While the theoretical parts could have been read at home, there was much more emphasis on that side in the lessons.

Therefore, since we had no experience in terms of practicing assessment, there will be problems in the future.’ (After.Low.Pre.4)

In addition, some of the pre-service teachers complained about not-updated content of the course. They mentioned that the samples were old and not suitable to recent teaching contexts. One of them indicated that:

‘Yes, the course content was sufficient. But the types [examples] in the coursebook we used in the lessons were a bit out of date. They may be updated. Apart from this, the course is beneficial and adequate.’ (After.Moderate.Pre.30)

Furthermore, one of the pre-service teachers drew attention to the mode of teaching. S/he criticized that course-takers did the presentations about the topics in the content most of the time, and thus, s/he wished the lecturer guided the lessons. S/he commented that:

‘The course partly responded to my needs. It was more efficient when the course lecturer was active in the class but for me, it was too hard to understand when the course takers did presentations about the topics.’ (After.Moderate.Pre.49)

Some pre-service teachers also mentioned that the semester of the course was too late to take because they believed they did not have any chance to practice their testing skills when they learned the knowledge of language assessment close to graduation. Similarly, they complained about the inadequate class hours of the course. Hence, one of them underscored that:

‘I think the course content was sufficient, but the problem is that the course is given too late [in the last semester]. If I had taken the course before, there would have been more chance to practice it in the last semester.’ (After.Moderate.Pre.95)

Thereupon, pre-service teachers made some recommendations to improve the course. They emphasized that the practical side should be increased by means of the assignments of preparing and administering language tests. Therefore, one advised that:

‘The course content was sufficient. However, the course takers would have been made to prepare exams according to the topics or to design mini quizzes in order to practice assessment.’ (After.Moderate.Pre.110)

Moreover, some of pre-service teachers recommended that there might be more examples related to integrated testing skills since they mentioned they learned testing each language skill separately. They also stated that more checklists can be given to learn how to evaluate and score. Besides, some suggested that the course might be coordinated

in line with the practicum where students may implement their own language testing tasks. To exemplify;

‘The course was sufficient in terms of its theoretical knowledge, but I think it was inadequate in terms of its practical side. I think the course should be integrated into the practicum.’
(After.Moderate.Pre.26)

As a consequence, pre-service teachers after the course appeared to appreciate language assessment more by realizing its importance and effect in teaching though most did not feel very qualified in FLAL especially due to the lack of practice and experience. In terms of their course experiences, although nearly all of them found the course beneficial and sufficient, they concentrated on some missing points of the course; and accordingly, made certain suggestions to develop it.

4.1.4. Findings from the documents about ELTE course content

Every ELT program at the undergraduate level in Turkey offers English language testing and evaluation course as the content and compulsory course. Since different course lecturers from different universities participated in this study, how their ELTE course was organized was examined by means of some public documents via the websites of the universities. There were 7 universities shared their opinions; so, the webpages about their course content were collected (Anadolu-ELTE Course, 2017; Bogazici-ELTE Course, 2017; Bursa Uludag-ELTE Course, 2017; ESOGU-ELTE Course, 2017; Hacettepe-ELTE Course, 2017; Marmara-ELTE Course, 2017; METU-ELTE Course, 2017) and analyzed by using the checklist constructed by the researcher beforehand. The detailed findings are given in Appendix-19.

To begin with, there are similar bits of information among the universities. For example, the title, the semester, the books used, and the assessment criteria of the course were very identical to each other with small differences. In terms of the books, though some universities preferred to use supplementary materials and books in the course, the most popular book belongs to Heaton. This book is about writing English language tests which focuses on the test development techniques for each language skill by giving practical examples (see Appendix-2 for the content list). As for the assessment of the course, most universities make use of midterm and final exams accompanied by some homework or assignments. Another similar finding is about the objectives of the course. Most universities highlighted the issue of theory which was acquired through the

knowledge of basic terms, principles and concepts of assessment, approaches to testing, types of tests, sample test items and techniques. They only focused on the practice by the design of appropriate language tests with regard to language skills. Similarly, with respect to learning outcomes, universities expect that their pre-service teachers both learn the rationale behind testing and assessment through studying the theories and are able to design appropriate language tests and interpret test scores. Presumably, the major finding is that lecturers designed and constructed the content according the book they used in the course. The documents showed that the main topics of the course were similar among the universities. They were about basic concepts of language testing, approaches to language testing, objective testing, types of tests, test preparation techniques for skills of English, practice of preparing various types of questions, evaluation and analysis techniques, and some statistical calculations (see Table 4.1 for the analysis summary).

Table 4.1. The main findings of the document analysis

<i>ELTE Course</i>	<i>Seven State Universities (Participant Course Lecturers' Universities) in Turkey</i>
Title	English Language Testing and Evaluation Language Assessment Foreign Language Measurement and Evaluation Assessment and Evaluation in English Language Teaching
Semester	7th or 8 th semester (3 hours a week)
Content of the Course/topic list	Introduction; Basic concepts of language testing; Approaches to language testing; Objective testing; Criteria (reliability, validity, etc.) and types of tests; tests of grammar, vocabulary, listening, speaking, reading and writing; interpreting test scores; analyzing techniques and statistical calculations; standardized tests
Objectives/ Learning outcomes	- basic terms and concepts of language testing, - basic principles of language testing, criteria of testing, approaches to language testing, - advantages and disadvantages of using published and class teacher-written tests, - evaluate advantages and disadvantages of test techniques, - engage in practices of language assessment, - design language tests appropriately, - prepare tests of language skills, and grammar and vocabulary, - write, implement, adapt and evaluate language tests, - interpret test scores and give feedback, - basic statistics.
Book(s) / Materials	Heaton, J. B. (2011). Writing English Language Test. Brown, H.D. (2003). Language assessment: principles and classroom practice. Hughes, A. (1989). Testing for language teachers. Genesee, F., & Upshur, J. A. (1996). Classroom-based evaluation in SLA. Bachman, L. (1990). Fundamental considerations in language testing.
Criteria for assessment and evaluation	Midterm: 40% (60%) Final: 60% (40%)

4.16). Each sample tended to give their general opinions about English language assessment and evaluation, and besides, they inclined to mention their own experiences about English language assessment. In terms of ELTE course itself (LAL training), all of them were aware of its importance and benefits while pointing to its deficiencies and how it can be improved more. Hence, it can be concluded that two themes of FLAL-Eng (Foreign Language Assessment Literacy in English) in this study can be named as: *Conceptual Dimension* (definitions, opinions, importance, general competencies, perceived competency) and *Training and Experiential Dimension* (ELTE course, its content and practices, its benefits and importance, its sufficiency and shortfalls, suggestions to improve; own use of ELTE in real life). In other words, in *Conceptual Dimension*, the questions of what they knew, what they needed to know, how much they were aware, and what kinds of changes there were in terms of competencies were answered. In *Training and Experiential Dimension*, the questions of what and how was practiced (topics, etc.), what was learned, what the missing points were, and what should be added or revised in terms of ELTE course were answered.

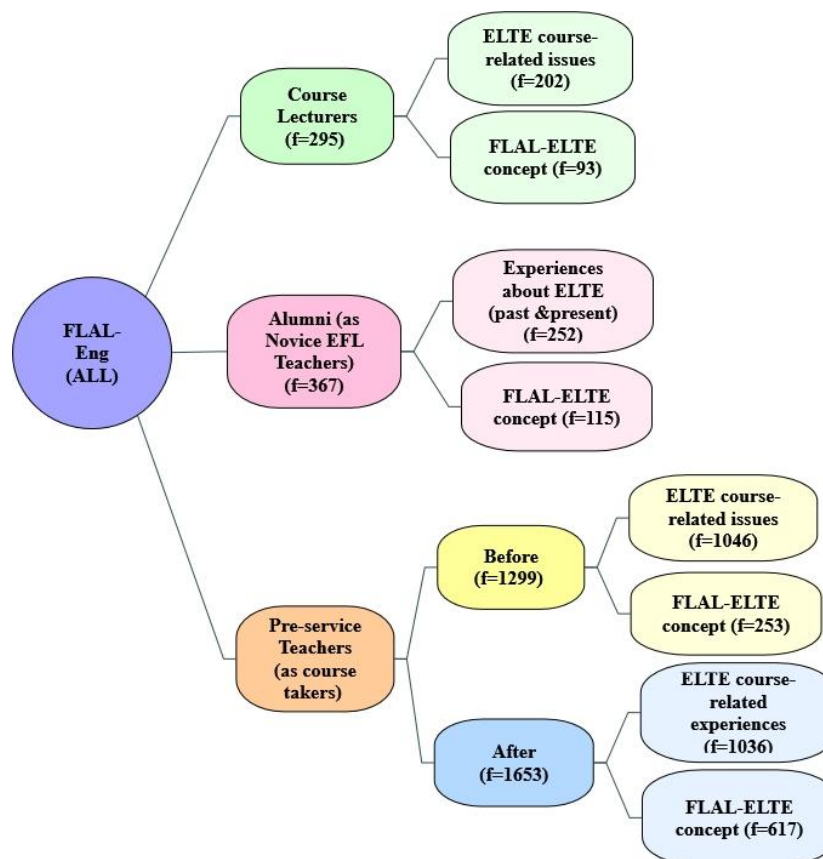


Figure 4.16. The thematic display of the findings for all the participants

To begin with the findings from the participant course lecturers, it can be seen that all of them began teaching ELTE course voluntarily or because the content interested them very much. Moreover, they thought that foreign language testing and assessment require both knowledge and experience. All the respondents made positive comments about the place of testing, assessment, and evaluation in ELT whereas some stated that only one course was not enough, for the content was too intense and demanding. Therefore, they suggested there needs to be elective courses or two following courses as I and II for a better education of foreign language assessment. Another finding is that most of them indicated that the course should be integrated into the practicum for more practice, and real-life experiences. Although some course lecturers tried to bring innovation to their courses by adding certain updated topics other than CoHE or the coursebook suggests such as Internet-based testing and assessing young learners, most agreed upon the adequacy of the course content, and they constructed their syllabi in line with the content of the book they used in the classroom. Furthermore, the criteria of evaluation of the students were the same for each course lecturer. Most respondents also reflected upon the product of the course: Their students seemed to be aware of the field more and good at traditional testing while students failed in implementation of testing because practice was missing. Lastly, all the course lecturers were aware of what foreign language assessment literacy mean, but some of their definitions focused only testing perspective.

Secondly, as far as the alumni graduated from Anadolu University within 5 years, that is, novice EFL teachers, are concerned, they perceived themselves good enough in foreign language assessment more or less, but they indicated they needed to improve themselves more because they stated they had difficulties in implementation of tests. Nearly all of them were also aware of the importance of the assessment and evaluation in foreign language teaching. As for the course itself, most of them found it sufficient and beneficial in terms of the content whereas some others discussed the problems regarding practice either external or internal factors; they suggested the practical side of the course should be highlighted. Most of them also recommended that the course content may be developed including some topics such as preparation of rubrics or checklist to score or control, statistical analysis and the interpretation and scoring of the results. Even they suggested more class hours required for such a kind of a course due to its rich content. Finally, as underlined by the graduates, it is a better idea to provide in-service training

about foreign language assessment owing to the need to remember what it is, and also the need for those who did not take this kind of course at their universities before.

Thirdly, all the pre-service ELT teachers who took the ELTE course in Spring Semester at Anadolu University were new to the course and senior students. Also, before starting the course, all the participants thought it is important and necessary to take the course because it would be beneficial for their teaching career, they would learn useful ways of detecting students' weaknesses and strengths, whether their teaching was effective, and also because the course would create awareness about testing and assessment preparation techniques. In addition, some of them indicated that the course would be different from the Turkish one due to its content of foreign language skills and areas testing. As for the expectations, they mostly emphasized the summative evaluation; to learn how to prepare language tests, to construct test items according to the level and skills, the criteria of testing and evaluation. Moreover, they expected to be a good test-writer, designer and evaluator. They highlighted the expectation of the practical side of the course; they wanted more practice rather than theoretical knowledge. Some of them, however, remarked that they expected to get a high score, and pass the course. About what they would learn, the comments were similar to their expectations; they hoped to learn how to design, score and evaluate language tests, the criteria for valid and reliable tests, and some statistical information. Finally, all of them believed they would benefit from the course because they indicated that they would learn to assess and evaluate students better, improve themselves, and realize the importance for their teaching profession.

Lastly, after pre-service ELT teachers took ELTE course, they still held the same opinion that assessment is important in the teaching/learning process. Their awareness of ELTE related issues increased, but they perceived themselves to handle language assessment at the moderate level. They mostly reported that though they gained theoretical knowledge of testing, they needed to improve their practical skills. In the same vein, although they found the course very beneficial and helpful, and the content sufficient, there were still missing points to be fulfilled, especially, in terms of practice. For example, they needed more samples or exercises related to each language skill testing as well as integrated testing to assess overall language ability. They also underlined the lack of actual practice of testing because only one course was not adequate to deliver all the information about assessment. Besides, they mentioned what they learned about

language assessment, and their abilities and capabilities gained after the course. Most of them considered themselves to be able to deal with language assessment practices though they needed more experience to do it. All in all, it seems that their views of testing, knowledge and skills of language assessment positively changed and improved after the course. However, their perception of assessment is based more on testing perspective most probably due to the topics and syllabus of the course content.

4.2.1. The explanation of the first research question

The present study seeks to answer, firstly, the question of what kind of LAL training needs pre-service ELT teachers may have. The analysis of all the opinions and experiences of the participant stakeholders (course lecturers, novice EFL teachers, pre-service teachers) and the documents about ELTE course content revealed that the course provided a basic training with pre-service teachers and offered the essential knowledge and theories of language testing. Most of the learning outcomes seemed to be achieved according to the reporting of the stakeholders in terms of the acquired knowledge and skills, and the changes and the competencies gained after the course. Yet, there have been still some gaps to be fulfilled to make teacher candidates of ELT be more competent in terms of LAL as well as be more effective teachers in their future careers. For example, on the basis of conceptual dimension, the perceptions of the stakeholders were oriented to testing notion rather than assessment and evaluation, and they considered assessment as something equals to measurement. There was also much more emphasis on the LAL dimensions as knowledge and skills than principles (proper and ethical use) and contexts (origins, reasons, and frameworks). Therefore, there is a need to change their perceptions into a more assessment perspective. That is, how assessment can be used to motivate and promote language learning apart from just measuring language proficiency. In addition, the topics of formative assessment, testing productive skills, interpretation and evaluation skills, statistics, integrated testing, Turkish contextual issues related to language testing, and ethical concerns were not much underlined in the course as mostly mentioned by pre-service and novice teachers. Hence, the stakeholders discussed that pre-service teachers were graduated with lack of such kind of knowledge and skills, and this argument was especially concentrated by novice EFL teachers as the difficulties in using their language assessment knowledge gained in the course in their teaching life. So, there is a need to enhance their knowledge and skills with respect to such points, as well. Nonetheless, the

most expected and needed area as highlighted and criticized by all the stakeholders was that the course was deprived of practice. Although it is good to see that pre-service teachers acquired theoretical knowledge and became more aware of language assessment issues, they did not have any opportunity to practice language testing and assessment in the course. Thus, they pointed out their practical skills did not much develop during the course period; they needed more practical exercises in the course and even, if possible, they needed and expected to experience assessment by constructing their own language tests and using some kind of formative assessment means in the practicum. It is because only teaching skills and methodological knowledge of pre-service teachers were assessed and evaluated through lesson plans and observations in the practicum; however, since testing/assessment and teaching are interrelated, some kind of assessment practices need to be also focused on in teaching practice to enable teacher candidates to appreciate the interaction between these two concepts as well as to put their knowledge into practice by experiencing it on their own. To sum up, in fact, the course helped pre-service teachers to develop their LAL at the basic level and most of the stakeholders believed their assessment skills could be improved through experience, but there have been still certain needs related to LAL that can be provided and focused through such training more before teacher candidates begin to teach and assess.

4.2.2. The explanation of the second research question

The current study examines, secondly, the effect of ELTE course on the LAL of pre-service ELT teachers. The analysis of all the data collected from three stakeholders and the documents indicated that the course was successful in providing fundamental language testing knowledge with pre-service teachers. Nearly all the stakeholders held the same opinion that the course was beneficial, and the content was sufficient in terms of language assessment training. For instance, course lecturers stated that as far as they observed, there were positive changes and most of the competencies about language testing were gained after the course. Similarly, both novice EFL teachers and pre-service teachers exemplified what they learned during the ELTE course. All the stakeholders highlighted that teacher candidates were more aware and critical of language testing issues and concerns, familiar with crucial concepts, principles, testing types and items, felt more competent, learned how to test each language skill and area separately for every language level and age group of learners. Besides, both pre-service and novice teachers

noted that most of their expectations and needs were met after they took ELTE course as the course-based training of LAL; that is, they were contented with the content much. Moreover, all the stakeholders underscored that it was good to take a specific English-focused language assessment course other than the Turkish one called ‘Measurement and Evaluation in Education’ because the Turkish one was very general and did not instruct how to test language skills. So, all the stakeholders stated that ELTE course was important for pre-service ELT teachers in that it prepared them for their future assessment practices in their language classes. Nevertheless, even though the course was useful to establish the basis for LAL training for future teachers of ELT, there have been certain training needs as indicated by the stakeholders as discussed in the first research question. For example, all the stakeholders admitted that the course contributed to theoretical knowledge more than practical knowledge; so, there was imbalance between theory and practice of language assessment during the course. In addition, the participants complained about the shortage of time due to the inadequate class hours and being a single course for LAL training because the content was very rich and demanding. Therefore, there was not much time to keep up with each planned goals of the course throughout the semester. Furthermore, considering the content; namely, the topics, it was based on the book (Heaton, 2011) used as the main coursebook, and thus, the topics were devoted to only language testing rather than assessment and evaluation. Likewise, when compared to CoHE’s (2007) suggested content and other universities’ topic list, the syllabus used at Anadolu University in ELTE course did not include some topics as also argued in the first research question. For instance, recent assessment means such as alternative assessment and computer-based testing, statistical procedures, analysis and interpretation skills such as scoring, marking and giving feedback, teacher-made vs ready-made tests, classroom-based vs large-scale/standardized tests, other forms of formative assessment, integrated testing for language skills and areas were not much mentioned in the ELTE course. All in all, it is apparent that though there was a positive change and effect in pre-service teachers regarding LAL after they took ELTE course, the course content needs revisions and updates in order to respond to the needs of future teachers with regard to their development of LAL. The last but not the least, it should be emphasized that to include such topics may not be sufficient; the knowledge should be practiced so as to make teacher candidates internalize what they learn, improve their LAL more and be ready for their future careers as more capable and effective teachers of English.

4.3. Discussion

The present study mainly aims to find out whether ELTE course as a course-based training in pre-service ELT education is effective to promote LAL of ELT teacher candidates as well as whether it is helpful to create changes about language assessment; that is, how LAL training in ELT training program through ELTE course was carried out. It also intends to reveal whether there are training needs about language assessment for further guidance. To address these issues, the opinions and experiences about language assessment were collected via open-ended questions from three stakeholders (course lecturers, novice EFL teachers and pre-service ELT teachers) and the documents about the ELTE course content were also gathered. The analysis of the data showed that there were two main inclinations as mentioned before: *Conceptual Dimension* (the theme *FLAL-ELTE Concept*) and *Training and Experiential Dimension* (the theme *ELTE Course-related Issues*). Under these dimensions, the participants discussed general opinions about ELTE itself and the issues related to ELTE course as LAL training in terms of its importance, sufficiency, shortfalls, and whether they are able to put what they learn into their teaching. The documents were also employed to explain and describe the content of ELTE better.

4.3.1. FLAL-ELTE concept

The concept of FLAL-ELTE was discussed from different points of views such as its definitions, general competencies, importance and own perceived competency.

In terms of the definitions, all the stakeholders similarly defined what ELTE means. Nearly all the stakeholders focused on the meaning of ELTE from the point of achievement/attainment and diagnostic purposes of testing. That is, they defined ELTE as the process of measuring and determining students' language level as their proficiency and their deficiencies by testing language skills and knowledge. Some of them also highlighted that ELTE is used to find out whether learning outcomes or course objectives are achieved while defining the construct of ELTE. Even though especially the course lecturers were aware of alternative assessment methods since they added such topics to their course, they did not mention such functions of assessment in their definitions. Thus, all the stakeholders did not differentiate the meanings of the terms testing, assessment, and evaluation. Besides, these explanations seem to define ELTE from a testing notion rather than assessment perspective because they emphasized summative evaluation rather

than formative evaluation. This result is in congruence with the findings of some research studies in which ELTE was considered only as summative assessment in the literature (e.g. Berry, Sheehan & Munro, 2017; Buyukkarci, 2014; Duboc, 2009; Gebril, 2017; Lam, 2014; Munoz, Palacio & Escobar, 2012; Oz, 2014; Saka, 2016). Giving priority to summative evaluation is most probably due the fact Turkey is an exam-oriented country which employs high-stakes testing in the form of traditional multiple-choice exams as discussed by especially novice EFL teachers and pre-service ELT teachers; and thus, the participants are more familiar with such testing types since they previously experienced. In addition, it may be stemmed from the fact that novice EFL teachers and pre-service ELT teachers are trained on the basis of a rather testing notion in their ELTE course seeing that the course content mainly concentrates on how to design tests and to write test items so as to measure language skills and areas. Therefore, there is a need to make teachers to comprehend the differences about the meanings of testing, assessment, and evaluation in the first place. Then, they should be aware of other functions of assessment such as how assessment can be used to improve learning/teaching performance and how it can be used to get and give feedback about teaching/learning. Also, it is important to make them to change their perspectives to keep with the recent conceptions of assessment (assessment for learning and formative assessment) as indicated by Fulcher (2012), Hidri (2018), Inbar-Lourie (2008), O'Loughlin (2006), Scarino (2013), and Tzagari and Vogt (2017).

As for the FLAL concept itself, the general competencies of FLAL that every EFL teacher should possess; that is, what is needed to be language assessment literate, were also stated. There are somewhat different opinions among the stakeholders. For example, the course lecturers defined FLAL as the understanding of language assessment concepts and principles, having the essential knowledge and skills of ELTE, the ability of developing and practicing assessment, and the ability to interpret test results, which are similar to the mostly discussed characteristics of LAL in the literature (e.g. Berry & O'Sullivan, 2016; Inbar-Lourie, 2013b; Newsfield, 2006; NTF, 2016; Rogier, 2014). They also underscored that FLAL requires not only knowledge but also experience as a distinct qualification which consists of being knowledgeable about the ELTE as well as being experienced in English testing both as test-taking and test-designing. On the other hand, pre-service teachers reported that FLAL means to be competent in how to measure and test knowledge of students about the target language and to know testing techniques. They also pointed out that every EFL teacher should possess this kind of competence

(FLAL) in order to assess whether students learn or not. Moreover, novice EFL teachers and some pre-service teachers noted that local needs or contextual factors such as school type or the general conception of testing in Turkey make EFL teachers be competent in carrying out appropriate assessment procedures in their classes, which requires them to be assessment literate in this sense - paying attention to the contextual needs in assessment. When compared to the literature regarding the definitions, dimensions, and characteristics of LAL, it can be inferred that the course lecturers were the most aware stakeholder about the meaning of FLAL, and they seemed to build idealized perceptions about FLAL competencies while not taking into account the contextual needs, the impacts, or historical, social and political frameworks as asserted by Fulcher (2012). Nevertheless, pre-service teachers focused more on the knowledge and practice side of FLAL whereas novice EFL teachers emphasized the contextual issues in assessment practices. Hence, it can be said that each stakeholder evaluated FLAL from different points of views. Regarding the dimensions of LAL as knowledge, skills, and principles as put forward by Davies (2008), it can be deduced that the dimension of *skills* (developing, applying and evaluating language tests) and the *knowledge* (how language and setting are described) were more highlighted by the stakeholders than the *principles* (the use of tests appropriately and ethically). Similarly, the stakeholders tended to think that the component of ‘*how*’ (implementation of language tests) and ‘*what*’ (descriptions of language to be measured) in FLAL were more important than the component of ‘*why*’ (the reasoning behind the choice of certain assessment methods than others) (Inbar-Lourie, 2008). In the same vein, according to Fulcher’s (2012) LAL dimensions, the *practices* (knowledge, skills and abilities to design, practice, and evaluate assessment) were more emphasized than the *principles* (the concepts and ethical concerns about assessment) and the *contexts* (origins, impacts, social, political and historical frameworks about assessment) by the stakeholders except for novice EFL teachers and some pre-service teachers. Furthermore, when compared to MoNE-ELT Competencies (2008) with respect to language assessment, it can be argued that the stakeholders commented more on the use of measurement and evaluation than the aims, interpretation, and feedback. Therefore, it can be concluded that most of the stakeholders underlined the competencies of FLAL as skills, practices and how it is used whereas they excluded the other features such as ethics, fairness, statistics to interpret, how to communicate with assessment results, knowledge of historical, social or political concepts related to assessment, the

rationale behind assessment choices and uses like contextual or instructional needs, and most importantly, how to describe the language aspects to be measured as stated by Fulcher (2012), Giraldo (2018), Inbar-Lourie (2008), and Stabler-Havener (2018). Thus, there is a need to draw attention to such features of FLAL, and to make them comprehend the meaning as well as the elements of being language assessment literate for better education.

However, when the perceived competency in FLAL of the participant stakeholders are considered, all the course lecturers and half of the novice EFL teachers considered themselves highly competent in English language assessment whereas pre-service ELT teachers after the course perceived themselves at the moderate level because they believed they would enhance their assessment skills via experience. Therefore, it can be interpreted that ELTE course made significant contributions to develop LAL of pre-service as well as novice EFL teachers because they believed they gained necessary knowledge and skills in assessing English language while those who felt less prepared to assess hoped to improve by means of experiencing assessment. Compared to the literature, although most of the studies found low levels of in-service or pre-service teachers (e.g. Buyukkarci, 2016; Fard & Tabatabaei, 2018; Karaman & Sahin, 2014; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014; Volante & Fazio, 2007; Y. Xu & Brown, 2017), the present study reached the conclusion that most of the stakeholders felt prepared to assess properly as in Hudaya's (2007) study. Thus, further training can be provided to make them gain practical skills more while experiencing English language testing and assessment.

Regarding the importance of FLAL-ELTE in education, all the stakeholders agreed that such knowledge and skills are very important in foreign language education. It is because as Bayat and Rezai (2015) asserted that there is a relationship between teaching and assessment, most of the participants mentioned that assessment is as important as teaching itself. For instance, most of the stakeholders reported that assessment is an effective feedback tool both for learning and teaching to detect the progress, the missing points, and whether students learn or not as supported by many researchers such as Bachman and Palmer (1996), Brennan (2015), Heaton (2011), Hidri (2018), and Rogier (2014). Besides, most of them, but especially pre-service teachers, underscored the importance of being language assessment literate because they pointed out that every EFL teacher should know how to assess as a part of their professional competence; so, they need to acquire the knowledge and skills of English language assessment. This idea has

been also justified by certain researchers similarly such as Buyukkarci (2016), Newsfield (2006), Rogier (2014) and White (2009), all of whom assume that better LAL leads to better education.

To sum up, all the stakeholders appreciated the importance and benefits of language assessment and being assessment literate in foreign language education. Although most of the stakeholders successfully defined and were familiar with the concept of ELTE and FLAL, they ignored certain features or competencies of FLAL as recommended by some researchers in the literature. It can be inferred that their knowledge was at the basic level because they focused on testing and summative functions of assessment, and the dimension of skills and practices mostly. Thus, there is a need to make them to be aware of various meanings and functions of language assessment while changing their beliefs and perceptions into more learning-oriented and formative practices of assessment.

4.3.2. ELTE course-related issues

The issues related to the training side of ELTE course, benefits and importance, shortfalls and suggestions, and the use of what was learned in teaching life as the experiential side were discussed.

4.3.2.1. *The content and practices of ELTE course*

The ELTE as a compulsory course is delivered to provide pre-service teachers with the knowledge and skills of English language testing and assessment at the last year of ELT training program. Though there is a suggested syllabus by CoHE as mentioned before, all the course lecturers mentioned that they were not aware of such content, and they organized their syllabus according to their own knowledge. However, when the documents, the statements of the course lecturers and the content recommended by CoHE are compared, it can be realized that the topics or the learning outcomes were quite similar. For example, most of the course lecturers demonstrated that their course comprised of the topics such as basic concepts and principles of language testing, approaches to language testing, different test types, testing techniques, and testing purposes, test developing techniques for different language skills and areas, criteria of testing like reliability and validity, and analyzing and interpreting test results. They also stressed that they added new topics such as alternative assessment, dynamic assessment, computer-based testing and statistics to their existing content. Thus, it appears that

although some thought that existing content was sufficient, others tried to update their content by adding new topics.

On the other hand, when compared to CoHE's content, it is apparent that some topics were not placed in their syllabi. For instance, classroom-based assessment vs standardized-tests, other types of assessment such as portfolio, self-assessment and learner diaries, and application of basic statistics were not covered much; only a few course lecturers tried to mention them in their courses. Thus, the topics give the impression that the syllabus is much more testing-oriented because of its highlighting testing issues as well as it seems to deal with mostly theoretical issues. This is also validated by novice EFL teachers and pre-service teachers in line with the opinions about the (in)sufficiency of the content. For example, most of them thought that the topics were very theoretical, and there was no time to practice their testing abilities; even, most of them stated they did not construct a language test on their own, and they were just assessed in terms of how much they knew about language testing. In addition, although most pre-service teachers stated that they expected more practice than theory, they pointed out they could not enhance their practical testing skills after the course. Hence, nearly all the pre-service and novice EFL teachers drew attention to the fact that the content was deprived of practical issues and it did not improve their testing language skills; they just became familiar with how to design tests, and what kind of test items can be used for each language skill testing. This argument was in fact expressed by the course lecturers, as well. They admitted the course content provided a basic training like introduction to language testing in terms of its theoretical side because they observed their course-takers became more aware, but they were not able to design and administer sound assessments. Such conclusions are also verified by some studies in the literature. For example, Jin (2010), Lam (2014) and Hatipoglu (2015) revealed that the course contained basic theoretical issues rather than practices and other topics about assessment. Therefore, some updates in terms of the content are needed, especially with regard to practical applications.

In addition, some novice EFL teachers argued that more emphasis was given to testing grammar, vocabulary and receptive skills than productive skills because they indicated they had difficulties in administering speaking and writing tests in their teaching professions. This finding was also found in Mede and Atay's (2017) study in which EFL teachers were not able to test listening and speaking. Similarly, most of in-service EFL

teachers were found to use language tests to measure grammar and vocabulary rather than testing other skills (e.g. Duboc, 2009; Saricoban, 2011; Semiz & Odabas, 2016; Tsagari & Vogt, 2017). Thus, there is a need to devote a balanced time among the topics of the course to create awareness as well as to provide good practices of all language skills testing for their future career.

Furthermore, some novice EFL teachers highlighted that some topics such as preparing rubrics or checklists should be added in order to improve their interpretation skills. Consequently, it can be inferred that the content did not provide such training much though the course lecturers stated this issue in their learning outcomes but not explicitly stated in their weekly planned topics in Anadolu-ELTE syllabus (2017). In this sense, Watanabe (2011) also highlighted that course-takers had challenges in marking and scoring after they took the course. Besides, most novice EFL teachers complained about the content did not reflect on Turkish contextual issues or its local needs seeing that they found hard to practice certain language testing types. Therefore, there is a need to develop interpretation skills, to emphasize and give examples related to local issues in the content. Then, the syllabus should be reorganized to include such topics in its content.

Pertaining to the materials used, all the course lecturers were found to use certain testing books; the most popular one was Heaton's (2011) 'Writing English Language Tests'. They chose their coursebook mostly according to the fact that the book contained practical samples. Yet, few of them included supplementary materials such as other books or articles, sample problems and questions, and statistical procedures. When the syllabus of the course lecturers is examined, it is clear that most of them arranged their topics according to the book content they used. Hence, the topics were based on much more testing perspective. Moreover, most pre-service teachers indicated that the book was outdated, and the examples did not match with recent assessment techniques in relation to the characteristics of language learners. These findings are similar to Davies's (2008) and Jin's (2010) studies in which most teachers adopted Heaton's book in their courses. However, it would be better if other supporting materials based on different kinds of assessment knowledge, methods and techniques, specifically, to show how assessment can be used with the aim of formative purposes to improve learning rather than just mentioning certain types of assessment were included.

Regarding the criteria for assessment and evaluation, it can be said that the criteria were the same for each course lecturer. They mostly employed a midterm and final exam

in the form of multiple-choice and open-ended questions. Nevertheless, some of them made students prepare language testing tasks, to present assessment topics and to write weekly reflections. Thus, it can be deduced that they also tested their course-takers in the form of summative assessment. This is also discussed by pre-service and novice EFL teachers such that they were assessed two times by a form of knowledge test mostly rather than preparing language testing tasks during the term. Thus, not just traditional but also other formative or alternative forms of assessment may be utilized to assess course-takers' progress, and this might also be an example for them to see how such assessment is carried out by experiencing it.

As a conclusion, it can be said that although it is good to see that the content was intended to improve testing abilities used in assessing English language skills and areas, and it gives a basic training how ELTE is implemented and the criteria of language testing such as reliability and validity, there is a need to make some revisions. It is because it gives the impression that the content was based on much more theoretical knowledge, and the topics were mostly about testing and summative assessment. In this way, course-takers developed a testing notion in their minds, and they focused on only certain purposes of assessment such as measuring language skills to find out how much is learned. Besides, compared to LAL dimensions and characteristics, it can be deduced that the content did not provide information about especially the component of principles (impact, ethics) (Davies, 2008), the rationale behind assessment choices (why) (Inbar-Lourie, 2008), the contexts (historical, social and political frameworks) as well as the principles and concepts (ethical issues) (Fulcher, 2012), and the characteristics of designing and executing language assessments in line with the learning goals. Thus, the syllabus can be redesigned in the light of pre-service teachers' needs as well as the contextual expectations and local needs, for a good training of LAL should include balanced proportions of what-how-why of language assessment on the basis of a contextualized instruction (Inbar-Lourie, 2008; Taylor, 2013).

4.3.2.2. The sufficiency and shortcomings of ELTE course

The experiences related to ELTE course illustrated that there were both positive and negative sides of the course in terms of its learning outcomes, the place in the curriculum, class hours, and the ability to use or difficulty in administering. Nonetheless, the findings

imply that in general, the course was found beneficial and effective, but some revisions and changes are needed to improve it more.

To begin with, all the stakeholders agreed that the ELTE course is important and beneficial because they believed it is helpful to provide necessary knowledge and skills about English language testing for the future professional life of teaching. It also enabled pre-service teachers and novice EFL teachers to become aware of and familiar with language testing issues. Most of them found the existing content was rich in supplying information how to test language skills and areas. Especially, pre-service teachers reported that ELTE course was complementary to teaching methods courses; so, they believed they made a connection between testing and teaching. They also underlined that ELTE course was a domain-oriented course directly related to testing English language abilities, which makes it be different from Turkish general education course of assessment. Besides, both pre-service and novice EFL teachers pointed out that the course responded to their needs and expectations. Moreover, there was a positive change before and after the course. For instance, the course lecturers stated that pre-service teachers were more aware of ELTE topics and gained basic knowledge of language testing. Novice EFL teachers also mentioned they were able to use the course knowledge in their classrooms to test their students properly, but mostly using traditional forms such as multiple-choice tests for grammatical ability. In the same vein, pre-service teachers underlined what the course contributed to them in terms of acquired knowledge and skills as well as gaining more critical views with respect to testing. Such findings are similar to those studies which concluded that ELTE course was beneficial in changing attitudes, gaining certain knowledge and skills, and achieving the course objectives (Csépes, 2013; C. Yetkin, 2015; Hilden & Frojdendahl, 2018; Yastibas & Takkac, 2018b; Watanabe, 2011).

Besides, some novice EFL teachers pointed out that it is important to take such course in teacher education for their teaching life because they discussed that some of their colleagues were not trained well in ELTE, and even there were those who did not take any training how to assess their students. This assertion is also verified by several studies in the literature such that most of the in-service EFL teachers either having basic training or having no training about ELTE were not able to handle language assessment in their classes properly; they stuck to traditional forms and their teaching and testing did not overlap (e.g. Gebril, 2017; Giraldo Aristizabal, 2018; Hakim, 2015; Koksall, 2004;

Oz, 2014; Vogt & Tsagari, 2014). In this sense, the participant novice EFL teachers suggested in-service training may be helpful to overcome such incompetencies while teaches are at the stage of teaching. They also pointed out in-service training would be helpful for those who already took course to refresh their knowledge and to learn from other teachers' experiences. The impact of in-service training that was helpful for teachers to gain assessment skills was also found in the literature (e.g. Gebril, Boraie & Arrigoni, 2016). Thus, in-service training can be said to be useful to fill the gaps in LAL of those who are not very competent even they took LAL training at teacher education as well as of those who are not familiar with and did not take such training before.

Furthermore, the course instructors were found very efficient and qualified in ELTE by pre-service teachers, which helps them to gain appropriate and beneficial knowledge of ELTE. This is similar to the argument of Jeong (2013) that the background of teacher educators is important such that they must be professionally developed to give content courses. Seeing that the participant course lecturers were experienced in giving ELTE course as well as open to innovations and made efforts to improve themselves, they were expected to be efficient teachers in ELTE.

As for the learning outcomes or the course objectives, the course lecturers pointed out that students would be more aware of ELTE, learn basic concepts, principles and types of ELTE, be able to develop, implement, adapt and evaluate language tests in terms of language skills and areas. Correspondingly, at the end of the course, most course lecturers indicated that their course-takers become more aware, familiar with basic concepts, critical about testing, acquire the ability to select and write appropriate test items, and feel more competent. When the perceptions of pre-service teachers are taken into consideration, they also pointed out that they believed their views of testing changed, they gained necessary knowledge and skills of ELTE in terms of its theories, learned criteria of testing (reliability, validity, table of specifications, etc.), types and purposes of tests, and how to construct appropriate test items according to related language skills and areas. In addition, most of novice EFL teachers reported that the course was useful to supply information about how to plan and apply language tests for each language skill and area. Therefore, it can be concluded that most of the course objectives were achieved, and most of their expectations and needs were met. The finding that the course was beneficial and effective was also confirmed by some studies in the literature, as well (e.g. Csépes, 2013; Hilden & Frojdendahl, 2018; Yastibas & Takkac, 2018b; Watanabe, 2011).

On the other hand, although the course was found beneficial, there were shortcomings highlighted by all the stakeholders. The most common shortfall was about the lack of practice during the course period. The course lecturers admitted that though their course-takers learned the theory of language testing, they did not develop their practicing skills of testing. Even though they stated that LAL entails both knowledge and experience, they did not include practical applications to make students to experience it. Besides, both pre-service teachers and novice teachers criticized that they learned only theory, but they could not transform their knowledge into practice. They also underlined that giving only one assignment as the preparation of language testing tasks did not mean to practice it. Therefore, they needed further training for practice as validated by certain studies in the literature (e.g. Buyukkarci, 2016; Hatipoglu, 2010; Jin, 2010; Komur, 2018; Sariyildiz, 2018; Sheehan & Munro, 2017; Viengsang, 2016; Wach, 2012).

Moreover, some pre-service teachers argued that they could not practice their ELTE knowledge in the practicum though they had a chance to do, but they said the mentors focused on teaching much rather than assessment since they were made to write lesson plans only on the basis of teaching. This argument is similar to some study findings such as Sariyildiz (2018) and Viengsang (2016) in which pre-service teachers were not able to practice assessment in the practicum. Thus, as many stakeholders in the present study suggested, teacher candidates must practice language testing not only in the ELTE course but also in the practicum for further improvement. In this issue, especially some novice EFL teachers mentioned their difficulties in testing their students with reference to the ELTE course. For example, they stated that they did not gain much experience during the course since they studied on imaginary situations too much. Thus, they found it hard to practice or adapt testing knowledge in real life. According to them, this was also stemmed from the fact that the course did not touch upon the topics related to contextual issues such local conditions or Turkish contextual issues though the context was very influential in testing practices (e.g. Saka, 2016; Yan, Zhang & Fan, 2018). Hence, they found the course a little bit imaginary and idealized in this sense. Compared to the literature, several studies concluded that most in-service teachers had the knowledge, but could not practice assessment procedures effectively as in the present study (e.g. Giraldo Aristizabal, 2018; Hakim, 2015; Jannati, 2015; Koksall, 2004; Mede & Atay, 2017; Oz & Atay, 2017; Yan, Zhang & Fan, 2018).

In addition, some novice EFL teachers and pre-service ELT teachers indicated that their interpretation skills and evaluation methods such as giving feedback, scoring and marking were not developed much. That is, they mentioned they had difficulties in communicating with assessment results. Though there were topics and learning outcomes related to analysis and interpretation of test results to get and give feedback in some of the syllabi of the universities, but not in Anadolu-ELTE (2017) syllabus, it appears that such issues were not covered in the course. Therefore, there is need to improve such skills in terms of language assessment seeing that most researchers have illustrated that one of the characteristics of being assessment literate is to be able to analyze and interpret the results to reinforce and motivate learning/teaching, and know how to communicate with results to give feedback (Berry & O'Sullivan, 2016; Inbar-Lourie, 2013b; NTF, 2016; Rogier, 2014; Stabler-Havener, 2018; Taylor, 2012). This characteristic of LAL is also included in MoNE-ELT Competencies (2008), and thus, teacher candidates need to develop such skills before starting to teach.

Concerning the semester, class hours, and the number of ELTE course, there are different opinions about them among the stakeholders. For instance, most of the course lecturers stated that to give ELTE in the last semester of teacher education is appropriate because pre-service teachers just finish their teaching methods courses and are ready to assess. Yet, some novice EFL teachers and pre-service teachers argued that the last semester was too late to take the course; it should have been given in earlier years of the program. This argument was also supported by one course lecturer, and s/he said ELTE can be added as a dimension to methods courses so that teaching and assessment are taught in an integrated way. In this way, pre-service teachers might be introduced to language testing and assessment earlier, and then, they may have a chance to improve themselves in ELTE course more. It is because all the stakeholders agreed that there must be a separate ELTE course in the program to help teacher candidates to be professionally competent in the field of language assessment. In addition, some of the course lecturers, novice teachers and pre-service teachers complained about the inadequacy of class hours because they noted that the class hours were not enough to cover such a comprehensive ELTE course-based training, and when LAL dimensions are considered, it can be inferred that all the components cannot be focused in a single course. They also commented that just a single course led the content to be lack of practical issues and there was no time to emphasize each topic in the same significance, and even some topics were not mentioned,

at all though they were in the syllabus. Hence, all the stakeholders who thought class hours and a single course were not adequate recommended that two following courses as ELTE-I and ELTE-II can be a solution; in the first one, more theoretical knowledge, and in the second one, more practical applications might be delivered. They also believed that elective courses may be helpful especially in relation to practice. However, it should be noted that elective courses do not guarantee that all teacher candidates select language assessment course as elective course; and hence, it is much more sensible to make the course compulsory. Seeing that the need for further training was stressed much in the literature as a result of the research studies (e.g. Hatipoglu, 2010; Sariyildiz, 2018; Sheehan & Munro, 2017; Tsagari & Vogt, 2017; Wach, 2012), there must be enough time for all the content of ELTE as well as the components of LAL in the course-based training to enable future teachers to be assessment literate.

In response to such shortfalls, all the stakeholders shared some suggestions to improve LAL training in ELTE course better. For example, a single course was not adequate to focus on all the topics of ELTE; and thus, they recommended that elective courses or dividing the course into two semesters may be helpful. Seeing that the most problematic side of the course was its lacking practical issues, some stakeholders suggested the ways to promote language testing practices; for instance, the coordination with practicum may be useful or more samples, examples, activities or tasks can be provided to improve practical skills. Specifically, since novice EFL teachers drew attention to the fact that testing productive skills were not much highlighted in the course, a balanced time and effort allocation to each topic can be beneficial. They also mentioned that more realistic examples and exercises may be helpful. More emphasis may be put on the tasks for interpretation and evaluation seeing as the content not covering such issues much. In addition, pre-service teachers underscored that they learned testing language skills separately, but they did not study on integrated testing in which all the language skills are tested to find out overall language ability as in proficiency and standardized-tests. Thus, the topic of integrated testing can be added to the content. Lastly, though nearly all the participants believed the course content was good and satisfying, they were all aware of the fact that it should be updated and revised.

To put it briefly, it can be concluded that the course was effective in giving basic LAL training because course-takers as novice and pre-service teachers agreed that they learned a lot and it created awareness about language assessment. Besides, more or less

the course objectives were achieved, and expectations and needs of course-takers were met. Still, there were some gaps to be filled. For instance, the perceptions of the stakeholders were based on a testing point of view than assessment, and mostly, traditional forms of testing methods as a summative assessment were studied in the course. It is because most of them highlighted that the purposes of testing and testing how much is learned by means of exams. Most of pre-service and novice EFL teachers perceived themselves to be able to carry out assessment procedures at the moderate level. Besides, not much emphasis was put on productive skills testing, interpretation and evaluation matters, and local needs and contextual issue. The integrated testing of skills was also not much underlined in the course. Specifically, the lack of practice and experience about language assessment was the most reported shortfall of the course by all the stakeholders. Hence, there is a need to revise the course content and practices in order to provide a better training of ELTE for future teachers and to make them more competent in LAL seeing that some of the characteristics to be highly competent in terms of LAL as discussed in the literature (e.g. Berry & O'Sullivan, 2016; Fulcher, 2012; Giraldo, 2018; Gotch & French, 2014; Huang & He, 2016; Stabler-Havener, 2018) were not fully provided and covered in ELTE course.

4.3.3. LAL training needs

The major concern of the current study is to explore foreign language assessment training needs of pre-service ELT teachers. It also aims to investigate to what extent a course-based training (ELTE course) is useful to improve their LAL before beginning their professional careers. The findings from three stakeholders concerned with the ELTE course at English language teacher education as well as the documents about the course content collected from other universities revealed the course provided the basic training for future teachers, there was a positive change in pre-service teachers after the course, and the course was found beneficial and helpful especially for its being English language focused. Yet, some shortcomings and challenges related to the course were highlighted by the participants, and also, some training needs for LAL were identified.

Firstly, the perceptions of the stakeholders regarding the concept of ELTE or FLAL were based on a testing notion rather than assessment perspective. It is because though they had certain knowledge of the basic terms and concepts of ELTE, they mainly focused on summative purposes of assessment which can be performed through traditional for

forms of language tests either to diagnose students' strengths and weaknesses or to measure their language knowledge and abilities. These findings were most probably caused by the topics covered in the course and the coursebook that was followed because the topics in line with the coursebook were much more testing-oriented. Therefore, there is a need to change their perceptions towards a more formative assessment viewpoint to catch up with the recent advancements and shifts in foreign language assessment and education as well as to gain the perspective of how to improve and foster foreign language learning via assessment rather than just measuring language on its own. Similarly, the stakeholders did not differentiate the meaning of testing, assessment and evaluation; they inclined to think all of them were the same, and they referred to testing and measurement. Thus, in this sense, there is also a need to make them be aware of what such terms mean. Considering the dimensions of LAL proposed in the literature (e.g. Davies, 2008; Fulcher, 2012; Inbar-Lourie, 2008), nearly all the stakeholders underlined the practices or skills which are related to how language assessment is conducted (to design, develop, write and administer language tests), and the knowledge that is about what language component is measured or assessed (grammar, vocabulary, listening, etc.). They did not much highlight the principles that are about how tests can be used properly and ethically, and also they did not much mention the contexts (typical frameworks for own teaching setting, origins, reason, impacts) and the reason why teachers decide to employ certain assessment methods. Therefore, there is a need to boost their LAL in terms of such dimensions to help pre-service teachers to be aware of what LAL means to develop their professional competence.

Secondly, concerning the objectives or learning outcomes of the ELTE course, the stakeholders shared their acquired knowledge and skills. They underlined that they became more aware and critical of ELTE, gained knowledge of basic concepts, became familiar with test types for different purposes and language levels, how to test language skills separately, and felt more competent. However, all the stakeholders underlined that the course was lack of practice at all points. For example, they criticized that though they learned how to design a grammar multiple-choice test item, they did not perform such knowledge during the course period. They also did not have any opportunity to use their testing knowledge in their teaching practice, but only their teaching skills were evaluated. Thus, assessment is separated from teaching even in the practicum, which is against the interconnection between assessment and teaching. Since they were not able to put their

knowledge into practice, there is a need to enable them to experience what they learn. Otherwise, as in the present study, all they learned remained at the theoretical base. Another point is that there was much more emphasis on testing grammar, vocabulary and receptive skills rather than productive skills; and thus, there is a need to allocate balanced proportions of times to each topic as they deserve. In addition to this, the participants indicated that integrated testing skills were not much highlighted; they only learned how to test each language skill on their own. It would be better if they studied and constructed an overall language test according to given aims. Furthermore, novice EFL teachers and some pre-service teachers argued that they did not have any training on how to score, interpret, give feedback and explain the results to others in order to enhance learning. So, since the interpretation and evaluation skills were not developed much, this issue should not be ignored. Moreover, contextual issues, local needs of Turkish educational system in relation to assessment and ethical concerns were not touched upon throughout the course. Therefore, there is a need to enable pre-service teachers to be aware of such issues. It can be inferred that some of the characteristics of assessment literate teachers were acquired, and the other features should be supplied to teacher candidates for their future teaching lives.

Thirdly, the content of ELTE course was mostly found sufficient by the stakeholders; however, they indicated that it was based more on theoretical knowledge than practical skills. When the content of ELTE course is compared with CoHE's (2007) suggested topic list for language assessment training and MoNE's (2008) expectations (see Table 4.2), it is obvious that the contents are similar to each other. Nevertheless, some dimensions or elements are not covered; and thus, they should be emphasized and included for a better LAL training. For example, some recent assessment methods such as alternative assessment (e.g. portfolio, self/peer assessment, diary) and computer-based testing, teacher-made assessments, statistical procedures, the analysis and interpretation of results to give feedback and testing overall language ability were not embraced much. As discussed previously, there was also no information about Turkish contextual testing issues that were highlighted much especially by novice EFL teachers. Then, there is a need to change testing-oriented content into more assessment-oriented by updating the syllabus as well as making it more practice-based.

Table 4.2. *The comparison and contrast of the ELTE course content*

	<i>CoHE suggests (2007)</i>	<i>Participant Universities (n=7) doing (2017-2018)</i>	<i>MoNE expects (2008)</i>	<i>What is needed? (theory and concept; findings of stakeholders and documents)</i>
1	basic concepts, principles and constructs of classroom-based assessment	introduction of the course: aim and course material, introduction to language testing; approaches to language testing	hidden in other competencies	✓ but LAL dimensions should be emphasized; also, more assessment perspective needed (especially formative)
2	different types of tests and testing and various types of questions for a wide range of language assessment purposes, development and evaluation of such language tests and of other available types	criteria and types of tests; objective testing no other assessment types such as alternative at Anadolu University but some mentioned other types such as alternative, computer-based, etc.	monitoring and assessing language development/ progress	✓ but more practice
3	language tests for different age groups, different proficiency levels and various learner styles	so-so	monitoring and assessing language development/ progress	✓ more examples needed
4	test preparation techniques for testing reading, writing, listening, speaking, vocabulary and grammar individually and testing language skills in an integrated manner	tests of grammar and usage; testing vocabulary, listening, oral production, reading, writing	monitoring and assessing language development/ progress	✓ but integrated manner and more practice needed
5	application of basic descriptive and inferential statistical calculations and the principles underlying test design	criteria and types of tests; no statistics at Anadolu University but some used statistics	none	statistics and interpretation needed with more practice

6	stages of test construction, item analysis and interpretation of test scores, standardized tests teacher-prepared language tests and beneficial backwash	stages of test construction in test preparation techniques for language skills & areas; backwash no interpretation and analysis techniques also, some mentioned standardized tests, interpretation of test results	monitoring and assessing language development/ progress	such topics need to be highlighted more, especially, interpretation of scores and teacher-made language tests needed with more practice
---	--	--	---	---

Lastly, as far as the procedural side of the course is taken into account, the criteria to assess and evaluate pre-service teachers were based on summative assessment since they were assessed two times by means of midterm and final exams. To set an example for future teachers of ELT, their progress in the course also needs to be assessed so that they become aware of other functions of assessment. Besides, most stakeholders complained about inadequate class hours and being a single course for ELTE; and thus, there is a need to redesign course hours and given semester. Moreover, the book used in the course was based on only language testing side. As a result, other supplementary materials need to be included for a better training for language assessment.

To sum up, the course seems to provide a basic training for LAL, and it would be better if it was revised in response to the needs of teacher candidates. In this way, it is hoped that they will be more aware, knowledgeable, skillful, and effective teachers in terms of language assessment.

CHAPTER 5

5. CONCLUSION

5.1. Summary of the Study

The current study is designed to investigate to what extent the course of English language assessment given in teacher education program at Anadolu University promotes the development of language assessment literacy of pre-service ELT teachers. It also aims to find out what kind of assessment training needs pre-service teachers have. To address these issues, this study, based on descriptive research, explored the opinions, experiences, needs, expectations and practices of three stakeholders (course lecturers, novice EFL teachers as the alumni, and pre-service teachers as the course-takers) through open-ended questions. The documents about the content of ELTE course applied in other universities were also collected by means of the related webpages on their university websites to support the findings while evaluating the course content. There were three groups of participants. The first group was made up of 10 course lecturers from 7 universities. The second group consisted of 22 novice EFL teachers as the alumni of ELT program at Anadolu University graduated within 5 years. Both course lecturers and novice EFL teachers responded to open-ended questions via a web-based application. The last group contained pre-service ELT teachers as the course-takers: Their responses to open-ended questions were gathered two times; before taking the course ($n=164$) and after they took the course ($n=146$). The responses of all the participants were analyzed via content analysis, and the documents about the course content were examined by the help of a categorical checklist. The analysis of the data revealed that nearly all the stakeholders' comments were similar and complementary to each other. Therefore, two themes for each stakeholder were found: *FLAL-ELTE Concept* which was about the conceptual dimension of LAL, and *ELTE Course-related Issues* which were about the training and experiential dimension of LAL. The participants defined the concept of FLAL and ELTE, underlined the importance of LAL, mentioned the qualifications and competencies of FLAL, and reported their own perceived competency in FLAL under the category of *FLAL-ELTE Concept*. Pertaining to *ELTE Course-related Issues*, each stakeholder shared their opinions and experiences from their own point of view. For instance, the course lecturers exemplified their course content and practices, commented on students' progress during the course, expressed why they taught that course, gave general opinions about the course

content, its topics, its offered semester, its length and its number. On the one hand, novice EFL teachers described their previous experiences in the course referring to acquired knowledge and skills, general benefits and sufficiency of the course, shortfalls of the course and suggestions to improve, and on the other hand, they focused on their current assessment experiences such as the ability to use what was learned in the course, how they used it, and the difficulties in performing assessment in their classes. Pre-service teachers before the course mentioned their expectations and needs and gave their general opinions about the course; for example, whether it was important and beneficial. After they took the course, pre-service teachers underscored what they acquired in terms of the knowledge and skills, whether their expectations were met, and discussed the course content with regard to its benefits and shortcomings.

Considering *FLAL-ELTE concept*, nearly all the stakeholders were aware of what ELTE means, familiar with some basic terms and concepts, and appreciated its importance in education. However, they concentrated on the testing notion, summative evaluation, and only the purposes of diagnosis and achievement in language testing. They did not much mention formative assessment or other types of testing and assessment. Therefore, they did not differentiate the meanings of testing, assessment and evaluation. Besides, all the stakeholders recognized that FLAL is an important feature of every EFL teachers' professional competence for better education, but in terms of the qualifications of FLAL, they mostly emphasized the dimensions of skills and knowledge rather than principles (proper and ethical use) and contexts (origins, reasons, historical, political and social frameworks). As for their perceived competency, all the course lecturers and half of the novice EFL teachers considered themselves to be highly competent in language assessment whereas pre-service teachers perceived themselves at the moderate level.

When it comes to *ELTE Course-related Issues*, in general, the course was found beneficial and helpful, especially for future teaching life. Moreover, the course content had similar topics and practices among the universities. There was also a positive change and improvement in terms of course-takers who felt themselves more competent after they took the course. They became familiar with basic terms and concepts, different test types and purposes, and how to test each language skill on its own. It can be inferred that it provided a basic training for LAL because all the stakeholders thought it was too theoretical, and there was no practice in the course. Besides, all the content was devoted to language testing; there were not much emphasis on other assessment topics such as

alternative assessment. Seeing that it was testing-oriented, the course-takers developed their testing knowledge and only learned how language skills can be tested with a great emphasis on testing grammar, vocabulary and receptive skills through traditional testing forms rather than productive skills and integrated testing without practicing their such knowledge. In addition, one of the topics criticized was about interpretation and evaluation skills which were ignored in the course; so, pre-service teachers did not know much how to score, grade or give feedback according to testing results. They also complained that the content and the book were outdated. Even, it did not touch upon local needs of Turkish EFL context as well as its education system, which shows the content ignored contextual issues. Furthermore, since the course was too demanding and there was no time to practice, the class hours were found inadequate. As a result, although the course seemed to be useful to promote LAL of pre-service teachers, there were still some missing points to be fulfilled in order to make teacher candidates more assessment literate; so, the ELTE course needs to be revised in response to the needs of pre-service ELT teachers for a better LAL training.

All in all, it can be concluded that this study is hoped to contribute to the understanding of LAL development and assessment training needs of pre-service ELT teachers within teacher education program.

5.2. Implications

The present study is believed to provide new insights into LAL development of pre-service ELT teachers by means of a course-based training at teacher training program, whether the training is helpful to enhance LAL, and what kind of assessment training needs teacher candidates have. The data collected by three stakeholders (ELTE course lecturers, novice EFL teachers, pre-service teachers) accompanied by the documents regarding the course content revealed positive sides and gaps of LAL training. Though it is very pleasing to find out the ELTE course was useful to supply initial basic LAL training, there are still some points to be revised to improve the training more in order to make pre-service teachers be ready for their future assessment duties. Therefore, in the light of the findings, some pedagogical implications can be suggested.

Language assessment courses can differ in terms of its purpose and target group (Berger, 2012); however, maybe their first objective is to make pre-service teachers be aware of what language assessment means (Coombe, Troudi & Al-Hamly, 2012). In the

present study, all the stakeholders were found to be aware of its meaning; however, they considered language assessment as something equals to language testing and measurement, and they did not differentiate the terms testing, assessment and evaluation. They also tended to think language assessment refers to summative rather than formative assessment. To broaden their perspectives about assessment, and to make them comprehend other functions of assessment such as how assessment contributes to learning/teaching (Jin, 2010; Klinger, 2016), the content of the course should be revised. It is because they shape their beliefs and perceptions about assessment according to what they learn, which affects their future assessment performance in their classes (Yan, Zhang & Fan, 2018). For example, at the beginning of the course, the different meanings of the basic terms and concepts should be explicitly introduced as also suggested by H. Xu (2017). In addition, other LAL dimensions such as principles, contexts, reasonings, origins, backwash effect and ethics should be emphasized because most of the stakeholders inclined to think only how testing procedure (developing, designing, writing test items) is executed rather than considering the reasons, purposes and contextual issues that influence assessment practices. In this respect, local needs of Turkish educational system can be discussed and exemplified throughout the lessons because each teaching context has its own concerns and challenges about assessment (Berger 2012). Unethical assessment practices should also be discussed to enhance assessment literacy (Newsfield, 2006) seeing that the stakeholders did not mention this dimension much. Furthermore, since the coursebook was testing-oriented, other topics about assessment and evaluation should be added to the syllabus. In other words, the syllabus should be updated by adding recent topics with regard to language assessment as done by some participant course lecturers in this study such as alternative assessment to enable them to understand how they improve language learning rather than only measuring language abilities as well as to help them to keep up with the recent developments. Besides, the course content was based on the testing language skills separately; however, pre-service teachers need to understand how an overall language ability is assessed. Therefore, the topic of the integrated testing and assessment should be included in the course. Moreover, a balanced time should be allocated to each topic in the course because some stakeholders complained about the greater emphasis put on testing grammar, vocabulary and receptive skills than productive skills; and thus, there is a need to focus on the ways of assessing speaking and writing skills. Also, most stakeholders criticized that they could not improve

their interpretation and evaluation skills to give feedback; so, this topic should be concentrated on more in the lessons. The evaluation criteria of the course itself should be changed because it is based on summative than formative; and hence, to be a good model for future teachers, other assessment criteria should be added to the course such as collecting drafts of language tasks designed by course-takers and evaluating them to improve their progress. Making the course a model and providing good practices are important to make pre-service teachers competent and assessment literate for their future assessment performances (Volante & Fazio, 2007).

On the other hand, not only adding needed topics to the syllabus and revising the content and objectives of ELTE course, but also promoting practicing skills about language assessment is needed. Otherwise, as in the present study, the course might remain too theoretical and without practice, pre-service teachers cannot develop meaningful assessment knowledge. To do this, more examples, exercises and tasks should be studied to make them to connect their knowledge with their skills of assessment. For instance, pre-service teachers can construct their own language tests with their evaluation criteria and with rationales written for their choices and preferences. Then, their sample tests can be discussed in the class in order to show how it can be improved more, whether the correct language component with appropriate type of testing or assessment is measured, why this type is chosen, whether the evaluation criteria are suitable to such samples, and so on. In this way, the whole class can learn from the experiences of their peers, and while they are giving feedback and reflecting upon their own tasks or tests, they can improve their critical views as well as consolidate their knowledge of assessment. In addition, ready-made or real tests and assessment tasks can be brought to the course in order to analyze and evaluate them to present and discuss well-written or poorly written items, and advantages and disadvantages using such items or tasks. Besides, their interpretation skills can be promoted by giving teacher candidates some samples to score and grade, and then, their such skills can be elaborated more through discussions and reflections. However, to feel real experiences of language assessment, there should be coordination between the practicum and the ELTE course (Sariyildiz, 2018; Viengsang, 2016). It is because teaching experience is a phase where pre-service teachers put their knowledge into practice. Thus, since assessment cannot be thought as a separate dimension from teaching, while writing lesson plans for the practicum, they may write appropriate testing tasks according to their learning outcomes and activities

and if it is convenient, they can perform them in the class. They might also write and implement other techniques of assessment suitably to their lessons such as questioning, giving feedback, checking understanding, peer-assessment and the like for the purposes of formative assessment in their weekly lessons in the practicum. Furthermore, they may be made to observe the class about assessment procedures, take notes and discuss them in the weekly lessons of the ELTE course. Moreover, one of the lesson plans written in the practicum can be based only on assessment, and such plans may be evaluated to figure out how pre-service teachers plan their assessment according to their teaching. In this way, the practicum may provide an opportunity for teacher candidates to gain experience in practicing their assessment skills in relation their teaching skills, as well.

Lastly, since the content of ELTE course is too demanding in terms of both theory and practice, there may be two following courses as ELTE-I (in the 7th semester) and ELTE-II (in the 8th semester) as suggested by the stakeholders. Since class hours increase in this way, there will be more time to focus on each topic as well as to practice in the course. In addition, pre-service teachers may be introduced language assessment in methods-classes with small examples or exercises to make them familiar with assessment earlier in the program, and then, they can develop their skills more while understanding the theories in the ELTE course.

In view of all the suggestions of the stakeholders and the implications, it is clear that the course content and practices should be revised and updated in response to LAL training needs of pre-service ELT teachers. Therefore, a model content can be recommended for a better course-based LAL training for teacher candidates of ELT. In this model, the course can be divided into two semesters seeing as the complaints about insufficient class hours and the need for more focus on the demanding content of language testing and assessment. In the first semester (the 7th semester of the ELT program), the theoretical background with examples about language assessment may be given, and in the second semester (the 8th semester of the ELT program), testing and assessing language skills and areas can be studied at length by means of doing practical exercises and if possible, performing language tests in the practicum; but if not possible, observing and reflecting on the assessment procedures of the classes where teaching practice is conducted. It should be highlighted that if language assessment were not practiced, pre-service teachers would not develop an understanding about the interconnection of assessment and teaching. The topics for the first semester (ELTE-I) may include the

introduction into language testing, assessment and evaluation, explicating and exemplifying their different meanings; the interaction between assessment and teaching; the introduction into what the concept of LAL refers to, what to be assessment literate in terms of professional teaching competence means, and how it can be helpful as to emphasize the rationale for taking ELTE course; the historical background of ELT assessment and how testing and assessment are viewed in Turkey; approaches to language testing with reference to recent developments in testing and assessment, and examples of Turkish implementations of language testing; the purposes, types and uses of multiple language tests and assessments such as comparing low-stakes vs high-stakes exams, formative vs summative assessment, classroom-based assessment vs standardized testing, recent forms like alternative assessment, dynamic assessment, and computer/Internet-based testing, commonly used question formats as well as the exercises on their examples, applications and ready-made and teacher-made measurement samples to comprehend how they can be designed with examples from Turkish testing system; the principles of language testing such as reliability, validity, washback and ethics accompanied by the discussion of their meanings, uses and examples; the preliminary introduction to language test development and the ways of analysis and interpretation of tests and assessments with sample exercises to study on by using basic descriptive statistics, and the revision of the semester. As for the topics for the second semester (ELTE-II), they might contain the revision of the first semester; test construction stages and the importance of table of specifications and how it can be formed; the ways of testing and assessing grammar and usage, vocabulary, reading comprehension skills, listening comprehension skills, oral production/speaking skills, writing skills/performance, integrated language skills-areas testing by studying on ready-made samples, doing exercises and activities, constructing own tests with rationales and evaluation criteria, discussing the practicum experiences and observations in that respect, reflecting on the possible performance of their own designed and used tests and assessment ways in teaching practice by means of administering and interpreting them; and the revision of the semester.

As a concluding remark, seeing that a good training of LAL should include balanced proportions of what-how-why of language assessment on the basis of a contextualized instruction (Inbar-Lourie, 2008; Taylor, 2013), such proposed revisions and the suggested model for language assessment courses according to the needs of pre-service ELT teachers might be helpful to promote their language assessment literacy.

5.3. Suggestions for Further Research

It can be assumed that the current study has contributed to the LAL research in the literature in many points such as revealing specific assessment training needs peculiar to ELT field at teacher education programs. For further study in line with the study findings, ELTE course can be revised, organized and coordinated with practicum in which some steps might be taken to observe whether the knowledge gained in the course is practiced effectively, and to investigate what can be done to deepen teacher candidates' language assessment skills more. In addition, seeing that this study focused on only one ELTE course as a sample and tried to describe how it was carried out to develop LAL, the experiences of other pre-service teachers from different universities can be included to elaborate more on the needs of them and thereby, to improve the quality of the ELTE course for a better LAL training. Likewise, only open-ended questions and documents were utilized to collect data, but other types of instruments such as questionnaires, interviews, observations and analysis of sample tasks created by course-takers can be employed to understand the LAL training more. Apart from descriptive research, other research designs may also be performed. For instance, in order to detect the development of LAL and immediately, respond to teacher candidates' needs to foster their assessment knowledge and skills during the course period, course lecturers can conduct action research to tailor their lessons accordingly. Moreover, comparisons between pre-service teachers and in-service teachers having different teaching experiences can be made to find out their practices and perceptions related to language assessment and its training to comprehend the concept of LAL better. Finally, a questionnaire or scale can be developed to investigate assessment training needs of pre-service ELT teachers according to course lecturers' own contexts to design the ELTE course content correspondingly.

REFERENCES

- Alderson, J. F. (2001). *Assessing reading*. Cambridge University Press.
- Anadolu-ELTE Course. (2017). *The description of English language testing and evaluation course of Anadolu University*. Retrieved from <https://www.anadolu.edu.tr/en/academics/faculties/course/114127/english-languagetesting-and-evaluation-ing/description> (Date of access: 29.10.2017).
- Bachman, L. F. (2005). *Statistical analyses for language assessment*. Cambridge University Press.
- Bachman, L. F. and Palmer, A.S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Baker, B. A. and Riches, C. (2017). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 0(0), 1-25. doi:10.1177/0265532217716732
- Bayat, K. and Rezaei, A. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1), 139-146.
- Bayram, E. (2015). *An investigation of language assessment preferences of pre-service EFL teachers and underlying factors* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/> (Date of access: 23.04.2017).
- Berger, A. (2012). Creating language-assessment literacy: A model for teacher education. In J. Hüttner, B. Mehlmauer-Larcher, S. Reichl & B. Schiftner (Eds.), *Theory and practice in EFL teacher education: Bridging the gap* (pp. 57-82). UK: Short Run Press Ltd.
- Berry, V. and O'Sullivan, B. (2016). *Assessment literacy for language teachers*. Retrieved from <http://tea.iatefl.org/wp-content/uploads/2015/10/Vivien-Berry-and-Barry-OSullivan.pdf> (Date of access: 04.05.2017).
- Berry, V., Sheehan, S. and Munro, S. (2017, May). *Exploring teachers' language assessment literacy: A social constructivist approach to understanding effective practices*. Paper presented at ALTE 6th International Conference Learning and Assessment: Making the Connections, Bologna, Italy. Retrieved from <http://eprints.hud.ac.uk/id/eprint/33342/> (Date of access: 16.12.2017).

- Bogazici-ELTE Course. (2017). *The description of English language testing and evaluation course of Bogazici University*. Retrieved from <https://fled.boun.edu.tr/undergraduate-course-descriptions> (Date of access: 29.10.2017).
- Bogdan, R. C. and Biklen, S. K. (2007). *Qualitative research for education: An introduction to theory and methods* (5th ed.). MA, US: Pearson Education Inc.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. doi:10.3316/QRJ0902027
- Brennan, M. (2015, May). *Building assessment literacy with teachers and students: New challenges?* Paper presented at ACER EPCC Conference, Sydney, Australia. Retrieved from <https://www.acer.org/files/eppc15-Brennan-Building-Assessment-Literacy-with-teachers-andstudents2.pptx+&cd=1&hl=tr&ct=clnk&gl=tr&client=firefox-b-d> (Date of access: 04.05.2017).
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Longman Education.
- Brown, J. D. and Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349-383. doi:10.1177/0265532208090157
- Buck, G. (2002). *Assessing listening*. Cambridge University Press.
- Bursa Uludag-ELTE Course. (2017). *The description of English language testing and evaluation course of Bursa Uludag University*. Retrieved from <http://bilgipaketi.uludag.edu.tr/Programlar/DetayENG/348?AyID=24> (Date of access: 28.04.2018).
- Buyukkarci, K. (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction*, 7(1), 107-120.
- Buyukkarci, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Egitim ve Ogretim Dergisi [Pegem Journal of Education and Instruction]*, 6(3), 333-346. doi:10.14527/pegegog.2016.017
- Campbell, C. and Duncan, G. (2007). From theory to practice: General trends in foreign language teaching methodology and their influence on language assessment. *Language and Linguistics Compass*, 1(6), 592-611. doi:10.1111/j.1749-818x.2007.00032.x

- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66. doi:10.1080/14703290601081332
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69(6), pp.963-976.
- Chen, P. P. (2005). Teacher candidates' literacy in assessment. *Academic Exchange Quarterly*, 9(3), 62-67. Retrieved from <https://www.thefreelibrary.com/Teacher+candidates%27+literacy+in+assessment.-a0138703662> (Date of access: 25.11.2017).
- Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, 20, 147-161.
- CoHE-ELTE Course. (2007). *The description of 'English language testing and evaluation' course by Council of Higher Education (undergraduate level-old program-compulsory)*. Retrieved from http://www.yok.gov.tr/web/guest/icerik//journal_content/56_INSTANCE_rEHF8BIsfYRx/10279/49875 (Date of access: 16.04.2017).
- Coombe, C. (2018). *An A to Z of second language assessment: How language teachers understand assessment concepts*. London, UK: British Council.
- Coombe, C., Troudi, S. and Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyloff (Eds.), *The Cambridge Guide to Second Language Assessment* (pp.20-29). NY, USA: Cambridge University Press.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). MA, US: Pearson Education Inc.
- Csépes, I. (2014). Language assessment literacy in English teacher training programmes in Hungary. In J. Hovarth & P. Medgyes, (Eds.), *Studies in Honour of Marianne Nikolov* (pp. 399-411). Pécs: Lingua Franca Csoport.
- Csépes, I. (2013, November). *Language assessment literacy in the Hungarian context: What is it like and how is it developed in English teacher training?* Paper presented at the 4th CBLA SIG Symposium, Cyprus. Retrieved from http://www.ealta.eu.org/events/CBLA-cyprus2013/Lectures_workshops/I_Csepes

- %20-%20%20pleanary%20-%20AL%20in%20Hungary.pdf (Date of access: 17.10.2017).
- Csépes, I. and Fekete, A. (2018a). Assessing reading skills. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis & S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 16-45). Nicosia, Cyprus.
- Csépes, I. and Fekete, A. (2018b). Assessing speaking skills. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis & S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 98-128). Nicosia, Cyprus.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347. doi:10.1177/0265532208090156
- Davison, C. and Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393-415.
- DeLuca, C. and Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438. doi:10.1080/0969594X.2010.516643
- DiBiase-Lubrano, M. J. (2018). The importance of assessment literacy for language faculty. *Teaching German [Die Unterrichtspraxis]*, 51(2), 121-131.
- Djoub, Z. (2017). Assessment literacy: Beyond teacher practice. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari & V. Thakur (Eds.), *Revisiting EFL assessment* (pp. 9-27). Switzerland: Springer International Publishing.
- Duboc, A. P. M. (2009). Language assessment and the new literacy studies. *Lenguaje*, 37(1), 159-178.
- Dunkley, D. (2015). Language assessment literacy in theory and practice. *愛知学院大学 語研紀要 [Foreign Languages & Literature]*, 40(1), 103-117.
- ESOGU-ELTE Course. (2017). *The description of English language testing and evaluation course of Eskisehir Osmangazi University*. Retrieved from [https://egitimiyabancidiller.ogu.edu.tr/Storage/egitimiyabancidiller/Uploads/InformationPackage-and-Course-Catalogue-\(2\).docx](https://egitimiyabancidiller.ogu.edu.tr/Storage/egitimiyabancidiller/Uploads/InformationPackage-and-Course-Catalogue-(2).docx) (Date of access: 29.10.2017).

- Fard, Z. R. and Tabatabaei, O. (2018). Investigating assessment literacy of EFL teachers in Iran. *Journal of Applied Linguistics and Language Research*, 5(3), 91-100.
- Fraenkel, J. R., Wallen, N. E. and Hyun, H. H. (2011). *How to design and evaluate research in education* (8th ed.). NY, US: The McGraw-Hill Companies Inc.
- Fulcher, G. (2010). *Practical language testing*. London, UK: Hodder Education.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. doi:10.1080/15434303.2011.642041
- Gay, L. R., Mills, G. E. and Airasian, P. W. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). NJ, US: Pearson Education Inc.
- Gebril, A. (2017). Language teachers' conceptions of assessment: An Egyptian perspective. *Teacher Development*, 21(1), 81-100. doi:10.1080/13664530.2016.1218364
- Gebril, A., Boraie, D. and Arrigoni, E. (2016). Assessment literacy training for English language educators in Egypt. In V. Aryadoust, & J. Fox (Eds.), *Trends in language assessment research and practice* (pp. 416-437). UK: Cambridge Scholars Publishing.
- Giraldo Aristizabal, F. G. (2018). A Diagnostic study on teachers' beliefs and practices in foreign language assessment. *Íkala: Revista de Lenguaje y Cultura*, 23(1), 25-44. doi:10.17533/udea.ikala.v23n01a04
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179-195. doi:10.15446/profile.v20n1.62089
- Giraldo, F. and Murcia, D. (2018). Language assessment literacy for pre-service teachers: Course expectations from different stakeholders. *Gist Education and Learning Research Journal*, 16, 56-77.
- Gotch, C. M. and French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2), 14-18.
- Gravetter, F. J. and Forzano, L-A., B. (2012). *Research methods for the behavioral sciences* (4th ed.). Belmont, CA: Wadsworth Cengage Learning.

- Green, A. (2016). Assessment Literacy for Language Teacher. In D. Tsagari (Ed.), *Classroom-based assessment in L2 contexts* (pp. 8-29). UK: Cambridge Scholars Publishing.
- Green, A. (2018). Assessing listening skills. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis & S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 81-97). Nicosia, Cyprus.
- Hacettepe-ELTE Course. (2017). *The description of English language testing and evaluation course of Hacettepe University*. Retrieved from [http://akts.hacettepe.edu.tr/ders_detay.php?ders_ref=DRSTNM_0000000000000000000000006486&ders_kod=%DDD%D6472&zs_link=1&prg_kod=449&submenueader=2&dil_kod=2](http://akts.hacettepe.edu.tr/ders_detay.php?ders_ref=DRSTNM_00000000000000000000006486&ders_kod=%DDD%D6472&zs_link=1&prg_kod=449&submenueader=2&dil_kod=2) (Date of access: 29.10.2017).
- Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education & Literacy Studies*, 3(4), 42-48. doi:10.7575/aiac.ijels.v.3n.4p.42
- Hamp-Lyons, L. (2017). Language assessment literacy for language learning-oriented assessment. *Papers in Language Testing and Assessment* 6(1), 88-111.
- Hamp-Lyons, L. (2018). Assessing writing skills. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis & S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 46-80). Nicosia, Cyprus.
- Hatipoglu, C. (2010). Summative evaluation of an English language testing and evaluation course for future English language teachers in Turkey. *English Language Teacher Education and Development (ELTED)*, 13, 40-51.
- Hatipoglu, C. (2015). English language testing and evaluation (ELTE) training in Turkey: expectations and needs of pre-service English language teachers. *ELT Research Journal*, 2015, 4(2), 111-128.
- Heaton, J. B. (2011). *Writing English language tests* (3rd ed.). UK: Longman Group.
- Herrera, L. and Macias, D. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. doi:10.14483/udistrital.jour.calj.2015.2.a09

- Hidri, S. (2018). Introduction: State of the art of assessing second language abilities. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice* (pp. 1-19). Cham, Switzerland: Springer International Publishing AG.
- Hilden, R. and Frojdendahl, B. (2018). The dawn of assessment literacy – exploring the conceptions of Finnish student teachers in foreign languages. *Apples – Journal of Applied Language Studies*, (12)1, 1–24. doi:10.17011/apples/urn.201802201542
- Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1), 1-17.
- Howerton, A. M. (2016). *Elephant on a stepladder: an exploration of pre-service English teacher assessment literacy* (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10240273) (Date of access: 25.04.2017).
- Huang, J. and He, Z. (2016). Exploring assessment literacy. *Higher Education of Social Science*, 11(2), 18-27. doi:10.3968/8727
- Hudaya, D. W. (2017). Teachers' assessment literacy in applying principles of language assessment. *Proceedings of Education and Language International Conference*, 1(1), 247-260.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. doi:10.1177/0265532208090158
- Inbar-Lourie, O. (2013a). Language assessment literacy. In C. A. Chapella (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-9). Blackwell Publishing Ltd. doi:10.1002/9781405198431.wbeal0605
- Inbar-Lourie, O. (2013b, November). *Language assessment literacy: What are the ingredients?* Paper presented at the 4th EALTA-CBLA SIG Symposium, Nicosia, Cyprus. Retrieved from http://www.ealta.eu.org/events/CBLA-cyprus2013/Lectures_workshops/O_Inbar-Lourie%20-%20plenary%20-1.pdf (Date of access: 24.04.2017).

- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, L. Or, & S. May (Eds.), *Language Testing and Assessment* (pp. 257-270). Springer International Publishing AG. doi:10.1007/978-3-319-02261-1_19
- 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. doi:10.1177/0265532213480334
- 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. doi:10.1177/0265532209351431
- Karagul, B. I., Yuksel, D. and Altay, M. (2017). Assessment and grading practices of EFL teachers in Turkey. *International Journal of Language Academy*, 5(5), 168-174. doi: 10.18033/ijla.3656
- Karaman, P. and Sahin, C. (2014). Ogretmen adaylarinin olcme degerlendirme okuryazarliklarinin belirlenmesi [Investigating the assessment literacy of teacher candidates]. *Ahi Evran Universitesi Kirsehir Egitim Fakultesi [Ahi Evran University Kirsehir Education Faculty Journal]*, 15(2), 175-189.
- Khadijeh, B. and Amir, R. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1), 139-146. doi:10.5296/ijele.v3i1.6887
- Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20(3), 77-102.
- Klinger, C. J. T. (2016). *EFL professors' beliefs of assessment practices in an EFL pre-service teacher training undergraduate program in Colombia* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10239483) (Date of access: 08.03.2018).
- Koh, E. T. and Owen, W. L. (2000). Descriptive research and qualitative research. In E. Koh & W. L. Owen (Eds), *Introduction to nutrition and health research* (pp. 219-248). Boston, MA: Springer.

- Koksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the Net. *Turkish Online Journal of Distance Education (TOJDE)*, 5(1), 1-11.
- Komur, S. (2018). Preservice English teachers' assessment awareness: Level of readiness for classroom practice. *The Journal of Language Teaching and Learning*, 8(1), 109-121.
- Kothari, C. R. (2004). *Research methodology: Methods and techniques* (revised 2nd ed.). New Age International (P) Ltd.
- Lam, R. (2014). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197. doi:10.1177/0265532214554321
- Lantolf, J. P. and Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research*, 15(1), 11-33. doi:10.1177/1362168810383328
- Looney, A., Cumming, J., van Der Kleij, F. and Harris, K. (2017). Reconceptualizing the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5) 1-26. doi:10.1080/0969594X.2016.1268090
- Luoma, S. (2009). *Assessing speaking* (5th ed.). Cambridge University Press.
- Malone, M. E. (2008). Training in language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *The Encyclopedia of Language and Education*, (2nd ed., Vol. 7) (pp. 225-239). Blackwell Publishing Ltd.
- Marmara-ELTE Course. (2017). *The description of English language testing and evaluation course of Marmara University*. Retrieved from http://dosya.marmara.edu.tr/aef/ydio/2015_sonras_.docx (Date of access: 26.03.2018).
- McKay, P. (2008). *Assessing young language learners* (3rd ed.). Cambridge University Press.
- Mede, E. and Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi-Ankara Universitesi TOMER [Language Journal-Ankara University TOMER]*, 168(1), 43-60.

- Mertler, C. A. (2003, October). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH, USA. Retrieved from the ERIC database (ED482277) (Date of access: 25.04.2017).
- METU-ELTE Course. (2017). *The description of English language testing and evaluation course of Middle East Technical University*. Retrieved from https://catalog.metu.edu.tr/course.php?course_code=4500413 (Date of access: 29.10.2017).
- Miles, M. B. and Huberman, A. M. (1994). *An expanded sourcebook: Qualitative data analysis* (2nd ed.). US: Sage Publications.
- Miles, M. B., Huberman, A. M. and Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). US: Sage Publications.
- Montee, M., Bach, A., Donovan, A. and Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 13(1), 1-31.
- MoNE-ELT Competencies. (2008). *İngilizce öğretmenleri özel alan yeterlikleri [The specialized field competencies of English language teachers]*. Retrieved from <http://otmg.meb.gov.tr/belgeler/alanyeterlikler/ingilizce%20%C3%B6%C4%9Fre%20tmenileri%20%C3%B6zel%20alan%20yeterlikleri.pdf> (Date of access: 01.06.2015).
- Munoz, A. P., Palacio, M. and Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile*, 14(1), 143-158.
- Newsfield, T. (2006). Teacher development and assessment literacy. *Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference*, 48-73.
- NTF (National Task Force on Assessment Education for Teachers). (2016). *Assessment literacy defined*. Retrieved from <http://michiganassessmentconsortium.org/wp-content/uploads/AssessmentLiteracyDefinedNationalTaskForce-May2016.pdf> (Date of access: 28.11.2017).
- O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student online subject forum. *University of Sydney Papers in TESOL*, 1, 71-85.

- Onalan, O. and Karagul, A. E. (2018). A study on Turkish EFL teachers' beliefs about assessment and its different uses in teaching English. *Journal of Language and Linguistic Studies*, 14(3), 190-201.
- Orr, M. (2010). Evaluating the testing course in an MA in ELT. *English Language Teacher Education and Development (ELTED)*, 13, 4-14.
- Oz, H. (2014). Turkish teachers' practices of assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research*, 5(4), 775-785. doi:10.4304/jltr.5.4.775-785
- Oz, S. and Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Ozarslan, O. O. (2017). Okuryazarlik kavrami [The concept of literacy]. In M. Yazici (Ed.), *Web okuryazarligi [Web literacy]* (pp. 3-26). Eskisehir, Turkey: Anadolu University Publication.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). US: Sage Publications.
- Pill, J. and Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402. doi:10.1177/0265532213480337
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(4), 4-11. doi: 10.1080/00405840802577536
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273, doi:10.1080/08878730.2011.605048
- Prasetyo, A. H. (2018). Language assessment literacy development: A student-teacher' experiences in teaching practice program. *LLT Journal: A Journal on Language and Language Teaching*, 21(2), 199-206. doi:10.24071/llt.2018.210209
- Purpura, J. (2005). *Assessing grammar*. Cambridge University Press.
- Razavipour, K. (2013). Assessing assessment literacy: Insights from a high-stakes test. *Research in Applied Linguistics*, 4(1), 111-131.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249-258. doi:10.1191=0265532204lt283ed
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

- Rogier, D. (2014). Assessment literacy: Building a base for better teaching and learning. *English Language Teaching Forum*, 3, 2-13.
- Sahinkarakas, S. (2012). The role of teaching experience on teachers' perceptions of language assessment. *Procedia – Social and Behavioral Sciences*, 47, 1787-1792. doi:10.1016/j.sbspro.2012.06.901
- Saka, F. O. (2016). What do teachers think about testing procedures at schools? *Procedia - Social and Behavioral Sciences*, 232, 575-582. doi:10.1016/j.sbspro.2016.10.079
- Saricoban, A. (2011). A study on the English language teachers' preparation of tests. *Hacettepe Egitim Fakultesi Dergisi [H. U. Journal of Education]*, 41, 398-410.
- Sariyildiz, G. (2018). *A study into language assessment literacy of pre-service English as a foreign language teachers in Turkish context* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/> (Date of access: 05.12.2018).
- 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. doi:10.1177/0265532213480128
- Scarino, A. (2017). Developing assessment literacy of teachers of languages: A conceptual and interpretive challenge. *Papers in Language Testing and Assessment*, 6(1), 18-40.
- Semiz, O. and Odabas, K. (2016). Turkish EFL teachers' familiarity of and perceived needs for language testing and assessment literacy. *Proceedings of the Third International Linguistics and Language Studies Conference*, 66-72.
- Sheehan, S. and Munro, S. (2017). Assessment: Attitudes, practices and needs. *ELT Research Papers*, 17.08. Retrieved from https://www.teachingenglish.org.uk/sites/teacheng/files/pub_G239_ELTRA_Sheenan%20and%20Munro_FINAL_web%20v2.pdf (Date of access: 25.04.2017).
- Spratt, M., Pulverness, A. and Williams, M. (2016). *The TKT-teaching knowledge test-course: Modules 1, 2, and 3* (2nd ed.). Cambridge University Press.
- Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Teachers College, Columbia University Working Papers in Applied Linguistics & TESOL*, 18(1), 1-22.

- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. doi:10.1017/S0267190509090035
- Taylor, L. (2012). *Developing assessment literacy*. University of Bedfordshire. Retrieved from http://in.beds.ac.uk/__data/assets/pdf_file/0010/197641/PROSET-Assessment-Literacy-Feb-2012.pdf (Date of access: 04.05.2017).
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. doi:10.1177/0265532213480338
- Tsagari, D. (2013). EFL students' perceptions of assessment in higher education. In D. Tsagari, S. Papadima-Sophocleous & S. Ioannou-Georgiou (Eds.), *International experiences in language testing and evaluation* (pp. 117-143). Frankfurt am Main, Germany: Peter Lang GmbH.
- Tsagari, D. and Vogt, T. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Turk, M. (2018). *Language assessment training level and perceived training needs of English language instructors: A mixed methods study* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/> (Date of access: 17.04.2019).
- Tuzcu-Eken, D. (2016). Testing in English classes in vocational schools of higher education. *Electronic Journal of Vocational Colleges*, 6(1), 38-47.
- Ukrayinska, O. (2018). Developing student teachers' classroom assessment literacy: The Ukrainian context. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice* (pp. 351-371). Cham, Switzerland: Springer International Publishing AG.
- Uyaniker, P. (2017). Language assessment: Now and then. *Avrasya Dil Egitimi ve Arastirmalar Dergisi [The Journal of Eurasian Language Education and Researches]*, 1(1), 1-20.
- Viengsang, R. (2016). Exploring pre-service English teachers' language assessment literacy. *Modern Journal of Language Teaching Methods (MJLTM)*, 6(5), 432-442.

- Vogt, K. and Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. doi:10.1080/15434303.2014.960046
- Vogt, K. and Froehlich, V. (2018a). Alternatives in assessment. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis & S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 148-178). Nicosia, Cyprus.
- Vogt, K. and Froehlich, V. (2018b). Providing feedback. In D. Tsagari, K. Vogt, V. Froehlich, I. Csépes, A. Fekete, A. Green, L. Hamp-Lyons, N. Sifakis & S. Kordia (Eds.), *Handbook of assessment for language teachers* (pp. 123-147). Nicosia, Cyprus.
- Volante, L. and Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770.
- Wach, A. (2012). Classroom-based language efficiency assessment: A challenge for EFL teachers. *Glottodidactica*, 39(1), 81-92.
- Watanabe, Y. (2011). Teaching a course in assessment literacy to test takers: Its rationale, procedure, content and effectiveness. *Cambridge ESOL: Research Notes*, 46, 29-34.
- Weigle, S. C. (2009). *Assessing writing* (6th ed.). Cambridge University Press.
- White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-assessment. *OnCUE Journal*, 3(1), 3-25.
- Xu, H. (2017). Exploring novice EFL teachers' classroom assessment literacy development: A three-year longitudinal study. *Asia-Pacific Educational Research*, 26(3-4), 219-226. doi:10.1007/s40299-017-0342-5
- Xu, Y. (2016). Assessment planning within the context of university English language teaching (ELT) in China: Implications for teacher assessment literacy. *Australian Review of Applied Linguistics*, 39(3), 233-254. doi:10.1075/aral.39.3.o2xu
- Xu, Y. and 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.

- Yan, X., Zhang, C. and Fan, J. J. (2018). “Assessment knowledge is important but ...”: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168. doi:10.1016/j.system.2018.03.003
- Yastibas, A. E. (2018). *Language assessment literacy of Turkish EFL instructors: A multiple-case study* (Doctoral dissertation). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/> (Date of access: 02.05.2018).
- Yastibas, A. E. and Takkac, M. (2018a). Understanding language assessment literacy: Developing language assessments. *Journal of Language and Linguistic Studies*, 14(1), 178-193.
- Yastibas, A. E. and Takkac, M. (2018b). Understanding the development of language assessment literacy. *Bingol University Journal of Social Sciences Institute*, 8(15), 89-106. doi:10.29029/busbed.364195
- Yetkin, C. (2015). *An investigation on ELT teacher candidates' assessment literacy* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/> (Date of access: 13.03.2018).
- Yetkin, R. (2017). *Pre-service English teachers' conception of assessment and their future assessment practices in a Turkish context* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/> (Date of access: 13.03.2018).
- Yin, R. K. (2011). *Qualitative research from start to finish*. NY: The Guilford Press.
- Zolfaghari, F. and Ahmadi, A. (2016). Assessment literacy components across subject matters. *Cogent Education*, 3, 1-16. doi:10.1080/2331186X.1252561

APPENDICES

- APPENDIX – 1:** The Syllabus of ELTE Course at ELT Department in Anadolu University
- APPENDIX – 2:** The Book Content of ELTE Course by Heaton (2011)
- APPENDIX – 3:** The Open-Ended Questions for Academics (ELTE Course Lecturers)
- APPENDIX – 4:** The Open-Ended Questions for Pre-Service ELT Teachers
- APPENDIX – 5:** The Open-Ended Questions for Alumni (Novice EFL Teachers)
- APPENDIX – 6:** The Sample of Consent Form
- APPENDIX – 7:** The Categorical Checklist for Document Analysis
- APPENDIX – 8:** The Research Ethics Approval
- APPENDIX – 9:** The Sample Screenshots of Google Forms
- APPENDIX – 10:** The Sample Screenshot of Responses from Google Forms
- APPENDIX – 11:** The Sample of the Paper given to Pre-Service Teachers
- APPENDIX – 12:** The Sample Screenshots of the Webpages about ELTE Course Content
- APPENDIX – 13:** The Data Analysis Performed by Hand (The First Impression)
- APPENDIX – 14:** The Code-Theme Relationship by NVivo (The Last Version)
- APPENDIX – 15:** The Findings from Course Lecturers (Detailed Version)
- APPENDIX – 16:** The Findings from Novice EFL Teachers (Detailed Version)
- APPENDIX – 17:** The Findings from Pre-service ELT Teachers (Before the Course) (Detailed Version)
- APPENDIX – 18:** The Findings from Pre-service ELT Teachers (After the Course) (Detailed Version)
- APPENDIX – 19:** The Findings of the Document Analysis
- APPENDIX – 20:** The Original Turkish Versions of the Quotes

APPENDIX – 1: The Syllabus of ELTE Course at ELT Department in Anadolu University

ANADOLU UNIVERSITY FACULTY OF EDUCATION ENGLISH LANGUAGE TEACHING DEPARTMENT

2017-2018 SPRING SEMESTER

NAME OF THE COURSE: İNÖ 401 English Language Testing

COURSE DESCRIPTION:

In this course, prospective teachers will have an understanding of the relation between teaching and testing and backwash effect of tests. They will also be introduced to different approaches to language testing and different types of tests. Besides, basic principles of testing such as validity and reliability will be studied. They will also learn the differences between objective and subjective tests. Through analysis of samples, they will learn how to test language areas (grammar and vocabulary) and language skills (listening comprehension, oral production, reading comprehension, and writing skills) and how to interpret these tests' scores. As future teachers, they will also practice evaluating and writing tests for these language areas and skills, calculating frequency of distribution of their own tests, conducting item analysis, and developing item cards and banks.

COURSE AIM

This course aims at raising awareness in testing and role of testing in language classes. The prospective teachers will also develop an understanding of basic principles of testing. Besides these, they will have a critical eye as future teachers for both analyzing and writing tests considering the needs of the target student group.

COURSE OUTLINE

Week 1	Introduction of the course: aim, course material, introduction to language testing
Week 2	Approaches to language testing
Week 3	Objective testing
Week 4	Criteria and types of tests
Week 5	Criteria and types of tests
Week 6	Activities
Week 7	Mid-term
Week 8	Tests of grammar and usage
Week 9	Testing vocabulary
Week 10	Listening comprehension tests
Week 11	Oral production tests
Week 12	Testing reading comprehension
Week 13	Testing the writing skills
Week 14	Revision

Course Requirements

Students are required to read the assigned materials and participate in class discussions actively and submit written assignments on time.

ASSESSMENT and EVALUATION

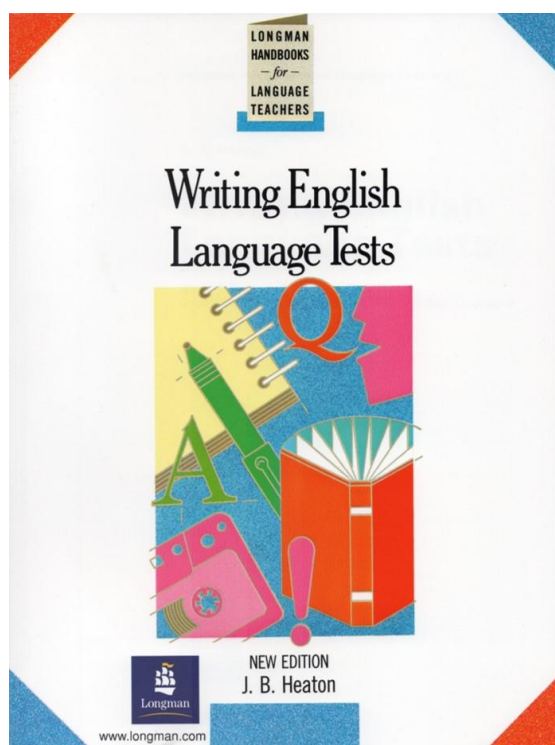
Midterm	30%
Assignment	30%
Final	40%

A note on assessment: Assessment for the mid-term will be an in-class exam. Assessment for the assignments will cover students' presentations on various chapters of the course-book. Students will develop tests for the language areas and skills covered in the lesson for the learner group identified by the instructor of the course. Assessment for the final will be another in-class exam.

COURSEBOOK

Heaton, J. B., 2011. Writing English Language Tests. Hong Kong: Longman.

APPENDIX – 2: The Book Content of ELTE Course by Heaton (2011)



Contents	
1 Introduction to language testing	5
1.1 Testing and teaching	5
1.2 Why test?	6
1.3 What should be tested and to what standard?	7
1.4 Testing the language skills	8
1.5 Testing language areas	9
1.6 Language skills and language elements	10
1.7 Recognition and production	11
1.8 Problems of sampling	12
1.9 Avoiding traps for the students	14
2 Approaches to language testing	15
2.1 Background	15
2.2 The essay-translation approach	15
2.3 The structuralist approach	15
2.4 The integrative approach	16
2.5 The communicative approach	19
3 Objective testing	25
3.1 Subjective and objective testing	25
3.2 Objective tests	26
3.3 Multiple-choice items: general	27
3.4 Multiple-choice items: the stem/ the correct option/ the distractors	30
3.5 Writing the test	33
4 Tests of grammar and usage	34
4.1 Introduction	34
4.2 Multiple-choice grammar items: item types	34
4.3 Constructing multiple-choice items	37
4.4 Constructing error-recognition multiple-choice items	39
4.5 Constructing rearrangement items	41
4.6 Constructing completion items	42
4.7 Constructing transformation items	46
4.8 Constructing items involving the changing of words	48
4.9 Constructing 'broken sentence' items	49
4.10 Constructing pairing and matching items	49
4.11 Constructing combination and addition items	50
5 Testing vocabulary	51
5.1 Selection of items	51
5.2 Multiple-choice items (A)	52
5.3 Multiple-choice items (B)	56
5.4 Sets (associated words)	58
5.5 Matching items	58
5.6 More objective items	60
5.7 Completion items	62
6 Listening comprehension tests	64
6.1 General	64
6.2 Phoneme discrimination tests	65
6.3 Tests of stress and intonation	68
6.4 Statements and dialogues	69
6.5 Testing comprehension through visual materials	71
6.6 Understanding talks and lectures	82
7 Oral production tests	88
7.1 Some difficulties in testing the speaking skills	88
7.2 Reading aloud	89
7.3 Conversational exchanges	90
7.4 Using pictures for assessing oral production	92
7.5 The oral interview	96
7.6 Some other techniques for oral examining	102
8 Testing reading comprehension	105
8.1 The nature of the reading skills	105
8.2 Initial stages of reading: matching tests	107
8.3 Intermediate and advanced stages of reading: matching tests	110
8.4 True/false reading tests	113
8.5 Multiple-choice items (A): short texts	116
8.6 Multiple-choice items (B): longer texts	117
8.7 Completion items	124
8.8 Rearrangement items	129
8.9 Cloze procedure	131
8.10 Open-ended and miscellaneous items	133
8.11 Cursory reading	133
9 Testing the writing skills	135
9.1 The writing skills	135
9.2 Testing composition writing	136
9.3 Setting the composition	138
9.4 Grading the composition	144
9.5 Treatment of written errors	149
9.6 Objective tests: mechanics	150
9.7 Objective tests: style and register	152
9.8 Controlled writing	154
10 Criteria and types of tests	159
10.1 Validity	159
10.2 Reliability	162
10.3 Reliability versus validity	164
10.4 Discrimination	165
10.5 Administration	167
10.6 Test instructions to the candidate	168
10.7 Backwash effects	170
10.8 Types of tests	171
11 Interpreting test scores	174
11.1 Frequency distribution	174
11.2 Measures of central tendency	175
11.3 Measures of dispersion	176
11.4 Item analysis	178
11.5 Moderating	185
11.6 Item cards and banks	185
Selected bibliography	188
Index	191

APPENDIX – 3: The Open-Ended Questions for Academics (ELTE Course Lecturers)

Türkçe Versiyonu	English Version
Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığı Çalışması	The Study on Foreign Language Assessment Literacy (FLAL)
1. Hangi üniversitede çalışmaktasınız? 2. Üniversitede kaç yıldır İngiliz Dili Eğitiminde Ölçme-Değerlendirme dersini vermektесiniz? 3. Bu dersi vermeye nasıl başladınız? 4. Yabancı dilde ölçme-değerlendirme alanındaki yeterlik düzeyinizi belirtiniz. () Yüksek () Orta () Düşük 5. Sizce bu dersi vermek ayrı bir yeterlik gerektiriyor mu? Lütfen düşüncelerinizi deneyimlerinizden de örnekler vererek belirtiniz. 6. Yabancı dilde ölçme-değerlendirme nedir? Kısaca tanımlayınız. 7. Ölçme-değerlendirmenin İngiliz Dili Eğitimi’ndeki yeri ve önemi hakkında neler söyleyebilirsiniz? 8. İngiliz Dili Eğitimi lisans programındaki İngiliz Dili Eğitiminde Ölçme-Değerlendirme dersi hakkında genel görüşleriniz üzerine; 8.1. Programda bu alanla ilgili tek ders olması hakkında neler düşünüyorsunuz? Lütfen nedenlerini deneyimlerinizden örnek vererek açıklayınız. 8.2. Sizce dersin lisans programında hangi yarıyılда verilmesi uygundur? Lütfen nedenleriyle açıklayınız. 8.3. Sizce ders içeriğinde hangi konular yer almalıdır? (YÖK’ten gelen programa ek olarak) 9. İngiliz Dili Eğitimi lisans programındaki İngiliz Dili Eğitiminde Ölçme-Değerlendirme dersi hakkında kişisel deneyim ve görüşleriniz üzerine;	1. At which university have you been working? 2. How long have you been teaching “English Language Testing and Evaluation (ELTE)” course? 3. How did you start teaching this course? 4. Indicate the qualification level in the field of foreign language assessment. () High () Moderate () Low 5. Does teaching this course require a distinct competence? Please express your opinions by giving examples form your experiences. 6. What is foreign language assessment and evaluation? Describe it briefly. 7. What can you say about the place and the importance of testing and evaluation in English Language Teaching (ELT) Education? 8. On your general opinions about the course of language assessment in the undergraduate program of ELT education; 8.1. What do you think about the undergraduate program having only one course in this field? Please express your opinions by giving examples form your experiences. 8.2. In which semester is it appropriate to place the course in the undergraduate program? Please explain it by providing reasons. 8.3. Which topics should be included in the content of the course? (In addition to

9.1. Verdiğiniz dersin izlencesinde hangi konular yer almaktadır? (Dersin izlencesini paylaşabilirsiniz) 9.2. Ders izlencenize YÖK’ün İngilizce Öğretmenliği Lisans Programı’nda bu dersle ilgili belirtilen konular dışında hangi konuları eklediniz? Neden? 9.3. Ders materyali ile ilgili; 9.3.1. hangi kitabı kullanıyorsunuz? 9.3.2. bu kitabı neye göre seçtiniz? 9.3.3. başka materyal kullanıyor musunuz? 9.4. Dersinizin değerlendirme ölçütleri nelerdir? 9.5. Dersinizin sonunda; 9.5.1. öğrencilerinizde bir değişim gözlemliyor musunuz? Ne/Nasıl? 9.5.2. onların yabancı dilde ölçme-değerlendirme yeterlilikleri hakkında neler söyleyebilirsiniz? 10. Sizce “Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığı” kavramını en iyi nasıl tanımlayabiliriz?	the program of Council of Higher Education-CHE) 9. On your personal experiences and opinions about the course of language assessment in the undergraduate program of ELT education; 9.1. Which topics are included in your course syllabus? (You can share your syllabus) 9.2. Which topics have you added to your syllabus except for the topics specified by the CHE related to this course? Why? 9.3. Related to the course material; 9.3.1. which book do you use? 9.3.2. according to which criteria did you choose this book? 9.3.3. do you use another material? 9.4. What are the criteria for the evaluation of your course? 9.5. At the end of your course; 9.5.1. do you observe any change in your students? What/How? 9.5.2. what can you say about their qualification level in the field of foreign language assessment? 10. How can you best describe the concept of “Foreign Language Assessment Literacy”?
--	---

APPENDIX – 4: The Open-Ended Questions for Pre-Service ELT Teachers

Türkçe Versiyonu	English Version
Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığı Çalışması	The Study on Foreign Language Assessment Literacy (FLAL)
Dönem Başı Soruları: 1. Anadolu Üniversitesi İngilizce Öğretmenliği Lisans Programında kaçınıcı sınıftasınız? 2. Programdaki İngilizce Sınav Hazırlama ve Değerlendirme (English Language Testing and Evaluation) dersini ilk defa mı alıyorsunuz? ☐ Evet ☐ Hayır 3. Bu dersin önemli ve verilmesinin gerekli olduğunu düşünüyor musunuz? Neden/hangi açıdan? 4. Bu dersten beklentileriniz nelerdir? 5. Bu dersin sonunda neler öğrenmiş olacağınızı düşünüyorsunuz? 6. Dersin faydalı olacağına inanıyor musunuz? Neden/hangi açıdan?	At the Beginning of the Term: 1. At which grade are you in the undergraduate program of English Language Teaching at Anadolu University? 2. Are you taking “English Language Testing and Evaluation” course in the program for the first time? ☐ Yes ☐ No 3. Do you think this course is important and it is necessary to lace it in the program? Why/in what way? 4. What are your expectations from this course? 5. What do you think about the things you are going to learn throughout the term? 6. Do you believe this course will be beneficial? Why/in what way?
Dönem Sonu Soruları: 1. Bu ders beklentilerinizi ve ihtiyaçlarınızı karşıladı mı? Nasıl/neden? 2. Ders sonunda neler öğrendiniz? Örnekler veriniz. 3. Ders faydalı oldu mu? 3.1. Evet: Hangi açıdan? 3.2. Hayır: Neden? 4. Ders içerik açısından yeterli miydi? Değilse eksiklikler nelerdi/neler olmalıydı? 5. Kendinizi yabancı dilde ölçme-değerlendirme alanında yeterli görüyor musunuz/hazır olduğunuzu düşünüyor musunuz? 5.1. Yeterlik düzeyi: ☐ Yüksek ☐ Orta ☐ Düşük 5.2. Neden/hangi anlamda? 6. Ölçme-değerlendirmenin İngiliz Dili Eğitimi’ndeki yeri ve önemi hakkında neler söyleyebilirsiniz?	At the End of the Term: 1. Did this course meet your expectations and needs? How/why? 2. What did you learn at the end of the course? Give examples. 3. Was the course beneficial? 3.1. Yes: In what way? 3.2. No: Why? 4. Was the course sufficient in terms of its content? If not, what were the shortcomings? 5. Do you think that you are qualified enough in/ready for foreign language assessment and evaluation? 5.1. Qualification level: ☐ High ☐ Moderate ☐ Low 5.2. Why/in what sense? 6. What can you say about the place and the importance of testing and evaluation in English Language Teaching (ELT) Education?

APPENDIX – 5: The Open-Ended Questions for Alumni (Novice EFL Teachers)

Türkçe Versiyonu	English Version
Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığı Çalışması	The Study on Foreign Language Assessment Literacy (FLAL)
1. Anadolu Üniversitesi İngilizce Öğretmenliği Lisans Programından mezun olduğunuz yıl: ☐ 2013 ☐ 2014 ☐ 2015 ☐ 2016 ☐ 2017 ☐ diğer 2. Şu an çalıştığınız kurumun; 2.1. adı: 2.2. seviyesi: ☐ Anaokulu / ☐ İlkokul / ☐ Ortaokul / ☐ Lise / ☐ Üniversite / ☐ Diğer 2.3. türü: ☐ Devlet / ☐ Özel 3. Yabancı dilde ölçme-değerlendirme alanındaki yeterlik düzeyinizi belirtiniz. ☐ Yüksek ☐ Orta ☐ Düşük 4. Yabancı dilde ölçme-değerlendirme nedir? Kısaca tanımlayınız. 5. Ölçme-değerlendirmenin İngiliz Dili Eğitimi'ndeki yeri ve önemi hakkında neler söyleyebilirsiniz? 6. İngiliz Dili Eğitimi lisans programındaki İngiliz Dili Eğitiminde Ölçme-Değerlendirme (English Language Testing and Evaluation) dersi hakkında; 6.1. Bu ders ihtiyaçlarınızı karşıladı mı? Hangi anlamda ve hangi alanda lütfen belirtiniz. 6.2. Aldığınız bu dersin içeriğini nasıl değerlendiriyorsunuz? Deneyimlerinizden örnekler vererek açıklayınız. 6.2.1. Ders içeriği yeterli değilse eksiklikler nelerdi? 6.2.2. Sizce ders içeriğinde neler olmalı? 6.3. Mezun olduğunuz lisans programında almış olduğunuz ölçme-değerlendirme derslerinin yeterlik düzeyini nasıl buluyorsunuz? Lütfen deneyimlerinizden örnek vererek açıklayınız.	1. The year you graduated from the Department of English Language Teaching at Anadolu University: ☐ 2013 ☐ 2014 ☐ 2015 ☐ 2016 ☐ 2017 ☐ other 2. The school you are currently working; 2.1. its name: 2.2. its grade: ☐ Kindergarten / ☐ Primary / ☐ Secondary / ☐ High School / ☐ University / ☐ Other 2.3. its type: ☐ State / ☐ Private 3. Indicate the qualification level in the field of foreign language assessment. ☐ High ☐ Moderate ☐ Low 4. What is foreign language assessment and evaluation? Describe it briefly. 5. What can you say about the place and the importance of testing and evaluation in English Language Teaching (ELT) Education? 6. About the course of language assessment (English Language Testing and Evaluation) in the undergraduate program of ELT education; 6.1. Did this course meet your needs? Please specify in what sense, and in which area. 6.2. How do you evaluate the content of this lesson you received? Explain it by giving examples from your experiences. 6.2.1. What were the shortcomings if the course content was insufficient? 6.2.2. What should be included in the course content? 6.3. What do you think about the sufficiency and efficiency of the courses about assessment you received in the

7. Şu an çalıştığınız kurumda bu dersin faydalarını gördünüz mü? Deneyimlerinizden örnek vererek açıklayınız. 8. Yabancı dilde ölçme-değerlendirme alanında hizmetiçi eğitim verilmesinin gerekli olduğunu düşünüyor musunuz? Neden?	undergraduate program you graduated from? Please explain it by giving examples from your experiences. 7. Have you benefited from this course at the school you are currently working? Explain it by giving examples from your experiences. 8. Do you think it is necessary to provide an in-service training program in the field of foreign language assessment and evaluation? Why?
--	--

APPENDIX – 6: The Sample of Consent Form

GÖNÜLLÜ KATILIM FORMU	CONSENT FORM
[ÜNİVERSİTE ÖĞRETİM ELEMANLARI; MEZUNLAR; ADAY ÖĞRETMENLER İÇİN]	[FOR ACADEMICIANS; ALUMNI; PRE-SERVICE TEACHERS]
Sayın/Değerli ...,	Dear ...,
Bu doktora tezi, İngiliz Dili Eğitimi öğretmen adaylarının yabancı dilde ölçme-değerlendirme okuryazarlığının geliştirilmesini ve İngilizce Öğretmenliği lisans programlarındaki İngiliz Dili Eğitiminde Ölçme-Değerlendirme dersinin buna etkisini araştırmayı amaçlamıştır.	This doctoral dissertation aimed to investigate the development of pre-service English language teachers' foreign language assessment literacy, and the impact of English Language Testing and Evaluation course in the undergraduate programs of English language teaching education on this literacy.
Bu çalışmaya katılımınız gönüllük esasına dayanmakta olup çalışmanın amacı doğrultusunda açık uçlu sorular yoluyla sizlerden veriler toplanacaktır. İsimleriniz ve paylaşmış olduğunuz bilgiler gizli tutulacaktır. Toplanan veriler bilimsel amaçlar doğrultusunda sadece bu çalışmada kullanılacaktır. Çalışmadan istediğiniz zamanda ayrılma hakkına sahipsiniz. Çalışmadan ayrıldığınız takdirde sizden toplanan veriler çalışmadan çıkarılacak ve imha edilecektir. Paylaşacağınız bilgiler araştırmanın güvenilirliği açısından önemlidir.	Your participation to this study is voluntary, and the data by means of open-ended questions in line with the purpose of the study will be collected from you. Your names and the information you shared will be kept confidential. The collected data will only be used in this study for research purposes. You have the right to leave the study at any time. If you leave, the collected data will be excluded and deleted. The information you share is important for the credibility of the research.
Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederim. Çalışma hakkındaki sorularınızı ilgili araştırmacıya yöneltebilirsiniz.	Thank you for your time to read and evaluate the consent form. You can ask your questions about the study to the concerned researcher.
Katılımınız ve katkılarınız için şimdiden teşekkür ederim.	Thank you in advance for your participation and contribution.

Saygılarımla, Arş. Gör. Aylin SEVİMEL ŞAHİN Anadolu Üniversitesi Eğitim Fakültesi Yabancı Diller Eğitimi Bölümü İngiliz Dili Eğitimi Anabilim Dalı Yunusemre Kampüsü Tepebaşı/ESKİŞEHİR Tel: +90-222-335-05-80 / 3498 E-posta: aylinsevimel@anadolu.edu.tr “Bu çalışmaya tamamen kendi rızamla, istediğim takdirde çalışmadan ayrılabilceğimi bilerek verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.”	Sincerely, Res. Assist. Aylin SEVİMEL ŞAHİN Anadolu University Faculty of Education Department of Foreign Language Education Program in English Language Teaching Yunusemre Campus Tepebasi/ESKISEHIR Tel: +90-222-335-05-80 / 3498 E-mail: aylinsevimel@anadolu.edu.tr “I voluntarily accept the information I shared to be used for research purposes while I know that I can leave the study when I wish.”
---	--

APPENDIX – 7: The Categorical Checklist for Document Analysis

1. Status of University
2. Name of University
3. Title of the Course
4. Type of the Course
5. Language of Instruction
6. Semester (Course Given)
7. Hours/ECTS of the Course
8. Mode of Delivery / Learning Activities & Teaching Methods
9. Instructor(s) of the Course
10. Book(s) Read in the Course
11. Content of the Course
12. Topic List of the Course
13. Assessment Criteria of the Course
14. Objective(s) of the Course
15. Learning Outcomes of the Course (Basic Key Words)

APPENDIX – 8: The Research Ethics Approval

Evrak Kayıt Tarihi: 13.12.2017	Protokol No: 139627	Tarih: 27.12.2017
--------------------------------	---------------------	-------------------



ANADOLU ÜNİVERSİTESİ
SOSYAL VE BEŞERÎ BİLİMLER BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU
KARAR BELGESİ

ÇALIŞMANIN TÜRÜ:	Doktora Tez Çalışması
KONU:	Eğitim Bilimleri
BAŞLIK:	Developing Foreign Language Assessment Literacy of Pre-service English Language Teachers (İngiliz Dili Öğretmen Adaylarının Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığının Geliştirilmesi)
PROJE/TEZ YÜRÜTÜCÜSÜ:	Yrd. Doç. Dr. Gonca SUBAŞI
TEZ YAZARI:	Aylin SEVİMEL ŞAHİN
ALT KOMİSYON GÖRÜŞÜ:	-
KARAR:	Olumlu

Prof.Dr. Coşkun BAYRAK
(Başkan-Eğitim Fak.)

Prof.Dr. T. Volkan YÜZER (Başkan Yardımcısı-Açıköğretim Fak.)	Prof.Dr. Esra CEYHAN (Eğitim Fak.)
Prof.Dr. Münevver ÇAKI (Güzel Sanatlar Fak.)	Prof.Dr. M. Erkan ÜYÜMEZ (İkt. ve İdari Bil. Fak.)
Prof.Dr. Handan DEVECİ (Eğitim Fak.)	Prof.Dr. Emel ŞIKLAR (İkt. ve İdari Bil. Fak.)

APPENDIX – 9: The Sample Screenshots of Google Forms

SORULARYANITLAR2

Bölüm 1/14

Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığı Çalışması

Sayın Öğretim Üyesi,

Bu doktora tezi, İngiliz Dili Eğitimi öğretmen adaylarının yabancı dilde ölçme-değerlendirme okuryazarlığının geliştirilmesini ve İngilizce Öğretmenliği lisans programlarındaki İngiliz Dili Eğitiminde Ölçme-Değerlendirme dersinin buna etkisini araştırmayı amaçlamıştır.

Bu çalışmaya katılımınız gönüllük esasına dayanmakta olup çalışmanın amacı doğrultusunda açık uçlu sorular yoluyla sizlerden veriler toplanacaktır. İsimleriniz ve paylaştığınız bilgiler gizli tutulacaktır. Toplanan veriler bilimsel amaçlar doğrultusunda sadece bu çalışmada kullanılacaktır. Çalışmadan istediğiniz zamanda ayrılma hakkına sahipsiniz. Çalışmadan ayrıldığınız takdirde sizden toplanan veriler çalışmadan çıkarılacak ve imha edilecektir. Paylaşacağınız bilgiler araştırmanın güvenilirliği açısından önemlidir.

Gönüllü katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederim. Çalışma hakkındaki sorularınızı ilgili araştırmacıya yönelebilirsiniz.

Katılımınız ve katkılarınız için şimdiden teşekkür ederim.

Saygılarımla,
Arş. Gör. Aylin SEVİMEL ŞAHİN

Anadolu Üniversitesi
Eğitim Fakültesi
Yabancı Diller Eğitimi Bölümü
İngiliz Dili Eğitimi Anabilim Dalı
Yunusemre Kampüsü
Tepebaşı/ESKİŞEHİR
Tel: +90-222-335-05-80 / 3498
E-posta: aylinsevimel@anadolu.edu.tr

Gönüllü Katılım Onay Formu

Açıklama (isteğe bağlı)

Bu çalışmaya tamamen kendi rızamla, istediğim takdirde çalışmadan ayrılabilceğimi bilerek verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.

☐ Kabul ediyorum.

Genel Bilgiler

Açıklama (isteğe bağlı)

1. Anadolu Üniversitesi İngilizce Öğretmenliği Lisans Programından mezun olduğunuz yıl: *

- ☐ 2013
- ☐ 2014
- ☐ 2015
- ☐ 2016
- ☐ 2017
- ☐ Diğer...

2.1. Şu an çalıştığınız kurumun adı: *

Kısa yanıt metni

2.2. Şu an çalıştığınız kurumun seviyesi: *

- ☐ Anaokulu
- ☐ İlkokul
- ☐ Ortaokul
- ☐ Lise
- ☐ Üniversite
- ☐ Diğer...

APPENDIX – 10: The Sample Screenshot of Responses from Google Forms

Otomatik Kaydet FIA Literacy (Academics) (Yanltar) - Excel Oturum açın

Dosya Giriş Ekle Sayfa Düzeni Formüller Veri Gözden Geçir Görünüm Ne yapmak istediğinizi söyleyin Paylaş

Yapıştır Pano Yazı Tipi Hizalama Sayı Stiller Hücreler Düzenleme

A	B	C	D	E	F	G	H
1 Zaman damgası	Bu çalışmaya tamamen k 1. Hangi üniversitede çalış 2. Üniversitede kaç yıldır 3. Bu ders vermede 4. Yabancı dilde ölçme-de 5. Sizce bu ders vermek ayrı bir ye 6. Yabancı dilde ölçme-değerlendir 7. Ölçme-değerlendirme						
2							
3							
4 11.7.2017 11:05:09	Kabul ediyorum.	Anadolu Üniversitesi	15	Kendi isteğimle	Yüksek	Alan ile bilgi sahibi olmak gerekli. Ç	Öğrencilerin dil becerileri konusu En az
5 11.7.2017 12:05:19	Kabul ediyorum.	Anadolu Üniversitesi	10	Ölçme Değerlendirme	Yüksek	Yabancı dilde ölçme-değerlendirme	Yabancı dil öğrenen kişilerin öğren
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							

Form Yanıtları 1

Hazır

Otomatik Kaydet FIA Literacy (Alumni) (Yanltar) - Excel Oturum açın

Dosya Giriş Ekle Sayfa Düzeni Formüller Veri Gözden Geçir Görünüm Ne yapmak istediğinizi söyleyin Paylaş

Yapıştır Pano Yazı Tipi Hizalama Sayı Stiller Hücreler Düzenleme

A	B	C	D	E	F	G	H	I
1 Zaman damgası	Bu çalışmaya tamamen k 1. Anadolu Üniversitesi İn 2.1. Şu an çalıştığınız kur 2.2. Şu an çalıştığınız kur 2.3 Şu an çalıştığınız kur 3. Yabancı dilde ölçme-de 4. Yabancı dilde ölçme-de 5. Ölçme-değerlendirme							
2								
3								
4 11.15.2017 17:05:12	Kabul ediyorum.	2016 n/a	n/a	Özel	Düşük	ogrencinin ogrenme surec Ölçme - değeri		
5 11.15.2017 21:31:05	Kabul ediyorum.	2017 Amerikan Kültür Derneği	Ortaokul, Lise, Üniversite	Özel	Orta	Dil öğrencilerin, dili iletişim Öğrencilerin iyi ya		
6 11.16.2017 14:35:36	Kabul ediyorum.	2017 Özel	Anaokulu	Özel	Düşük	Öğrencinin neyi biliş biliş Her alanda olduğu		
7 11.16.2017 14:41:33	Kabul ediyorum.	2017 American Culture Language	Lise, Üniversite	Özel	Yüksek	Kendini düzgünce ifade et Eger değeri		
8 11.16.2017 16:03:41	Kabul ediyorum.	2017 Altınbaş Üniversitesi	Üniversite	Özel	Orta	Öğrencilere sunulan bilgisi Bu alanda önemli t		
9 11.16.2017 18:03:34	Kabul ediyorum.	2017 Hava harp okulu	Üniversite	Devlet	Orta	Öğrencilerin bildiklerini gö Ölçme değeri		
10 11.16.2017 23:23:36	Kabul ediyorum.	2017 Teol dil okulları	İlkokul	Özel	Yüksek	Ders kapsamında kazandı Belirlenen kazanım		
11 11.18.2017 10:01:18	Kabul ediyorum.	2017 75. Yıl Çok Programlı Anı	Lise	Devlet	Yüksek	Dil yeterliliğini belirli bir ö Her konuda olduğu		
12 11.18.2017 18:35:39	Kabul ediyorum.	2015 75. Yıl İlkokulu	İlkokul	Devlet	Orta	yabancı dilediği yeterliliği Ölçme-değeri		
13 11.21.2017 11:34:23	Kabul ediyorum.	2017 Meb	Ortaokul	Devlet	Orta	Öğrencilerin dili kullanmaları		
14 11.24.2017 9:21:42	Kabul ediyorum.	2017 Dil okulu	Her seviye	Özel	Orta	Yapılan sınavların doğru b Dil öğretiminin dog		
15								
16								
17								
18								
19								
20								
21								

Form Yanıtları 1

Hazır

APPENDIX – 11: The Sample of the Paper given to Pre-Service Teachers

YABANCI DİLDE ÖLÇME-DEĞERLENDİRME OKURYAZARLIĞI ÇALIŞMASI

Görüşme Katılım Formu

Değerli Öğrenci,

Bu doküman test, İngiliz Dili Eğitimi öğretmen adaylarının yabancı dilde ölçme-değerlendirme okuryazarlığının gelişimini ve İngilizce Öğretmenliği lisans programlarındaki İngiliz Dili Eğitiminde Ölçme-Değerlendirme dersinin buna etkisini araştırmayı amaçlamıştır.

Bu çalışmaya katılımını gönüllük esasına dayandırmakta olup çalışmanın amacı doğrultusunda açık uçlu sorular yoluyla sizlerden veriler toplanacaktır. İstifliminiz ve paylaşımlığınız bilgiler gizli tutulacaktır. Toplanan veriler bilimsel amaçlar doğrultusunda sadece bu çalışmada kullanılacaktır. Çalışmadan istediğiniz zamanla ayrılmaya hakkına sahipsiniz. Çalışmadan ayrıldığınız takdirde sizden toplanan veriler çalışmadan çıkarılacak ve mihna edilecektir. Payışacağınız bilgiler araştırmanın güvenilirliği açısından önemlidir.

Görüşme katılım formunu okumak ve değerlendirmek üzere ayırdığınız zaman için teşekkür ederim. Çalışma hakkındaki sorularınızı ilgili araştırmacıya yöneltebilirsiniz.

Katılımınız ve katkılarınız için şimdiden teşekkür ederim.

Saygılarımla,
Arş. Gör. Aylin SEVİMEL ŞAHİN

Anadolu Üniversitesi
Eğitim Fakültesi
Yabancı Diller Eğitimi Bölümü
İngiliz Dili Eğitimi Anabilim Dalı
Yunusemre Kampüsü
Tepebaşı/ESKİŞEHİR
Tel: +90-222-335-05-80 / 7498
E-posta: aylinsevime@anadolu.edu.tr

"Bu çalışmaya tamamen kendi rızamla, istediğim takdirde çalışmadan ayrılabileceğimi bilerek verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum."

Yabancı Dilde Ölçme-Değerlendirme Okuryazarlığı Çalışması

Dönem Sonu Soruları:

1. Anadolu Üniversitesi İngilizce Öğretmenliği Lisans Programında kaçınıcı sınıfınız?
4

2. Programdaki İngilizce Sınav Hazırlama ve Değerlendirme (English Language Testing and Evaluation) dersini ilk defa mı alıyorsunuz?
☒ Evet ☐ Hayır

3. Bu ders beklentilerinizi ve ihtiyaçlarınızı karşıladı mı? Nasil/neden?
Dersi almadan önce ders hakkında birşey bilmiyordum. Ders beklentilerimi bir hayli karşıladı çünkü nasıl sınav hazırlanır, neyi ne kadar çalıştırır, konularda bilgiler alınırmı, bir çalışmada sahip olma gerektir ve böylece en önemli şeyler bir sınav hazırlanırken önemli olan sadece öğrenmek değil aynı zamanda olmaktadır. Bu ders sayesinde bu doğru bir şekilde yapılabileceğine inanıyorum.

4. Ders soruları neler öğrendiniz? Örnekler veriniz.
"Dersler esnasında her skille ne kadar sonra yapılıyor ve bununla ilgili ne kadar soru sorulmalı sorulmalı?" Bu soruların cevabını öğrendim. Table of specifications tabiriyle neyi ne kadar çalıştırmanın etkiler, sorulara cevapların, Reliability and validity önemli kavramlarıdır ki bir sınavda bu iki özellik yoksa o sınav doğru hazırlanıp uygulanamaz denir.

5. Ders faydalı oldu mu?
5.1. Evet. Hangi açıdan?
Herde ders anlatılmamış sahneler, yeni öğrenen soru sorulmuş ve nasıl ne kadar soru sorulmuş, nasıl hazırlanmalı ve nasıl değerlendirilmeliye konusunda yardımcı oldu.

5.2. Hayır. Neden?
Hayır.

6. Ders içerik açısından yeterli miydi? Değilse eksiklikler nelerdir/neler olmalıydı?
İçerik açısından yeterli olduğunu düşünüyorum. Hazırlanan ders kitabında yeterince içerik yoktu. İçerik eksiklikleri vardı ve ders boyunca ile konularla derinleşmeye inceleştik.

7. Kızdığınız, yabancı dilde ölçme-değerlendirme alanında yeterli görüşer bulunmaz/bazı olduğunuzu düşünürsünüz mü?
7.1. Yeterlik düzeyi: ☒ Yüksek ☐ Orta ☐ Düşük

7.2. Neden/hangi anlamda?
Yeterli düşünüyorum çünkü sınav hazırlama gereklilikleri hakkında bilgime düşüncelerim Reliability ve validity için sorular hazırlanarak zorunlu olarak bu konuda yeterli bilgim var ve bunlara uygun sorular hazırlayacağım.

8. Ölçme-değerlendirme İngiliz Dili Eğitimi'ndeki yeri ve önemi hakkında neler söyleyebilirsiniz?
Her öğrencinin bilmesi ve doğru bir şekilde uygulanması gereken bir şey. İngiliz Dili'ni öğrenen bir öğrencinin en önemli şeyler, doğru bir şekilde öğrenmesi ve doğru şekilde uygulanması. Çok iyi değerlendirilirse ve doğru şekilde uygulanırsa süreçte bunun pek bir önemi yoktur.

APPENDIX – 12: The Sample Screenshots of the Webpages about ELTE Course Content

- Description
- Content
- Learning Outcomes
- Learning Activities and Teaching Methods
- Course's Contribution to Prog.
- Assessment Methods

Description

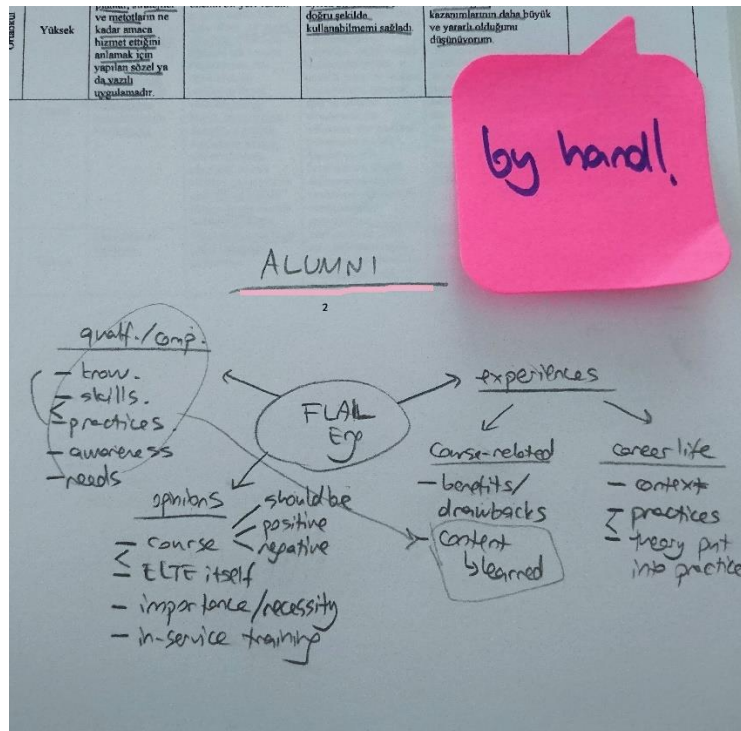
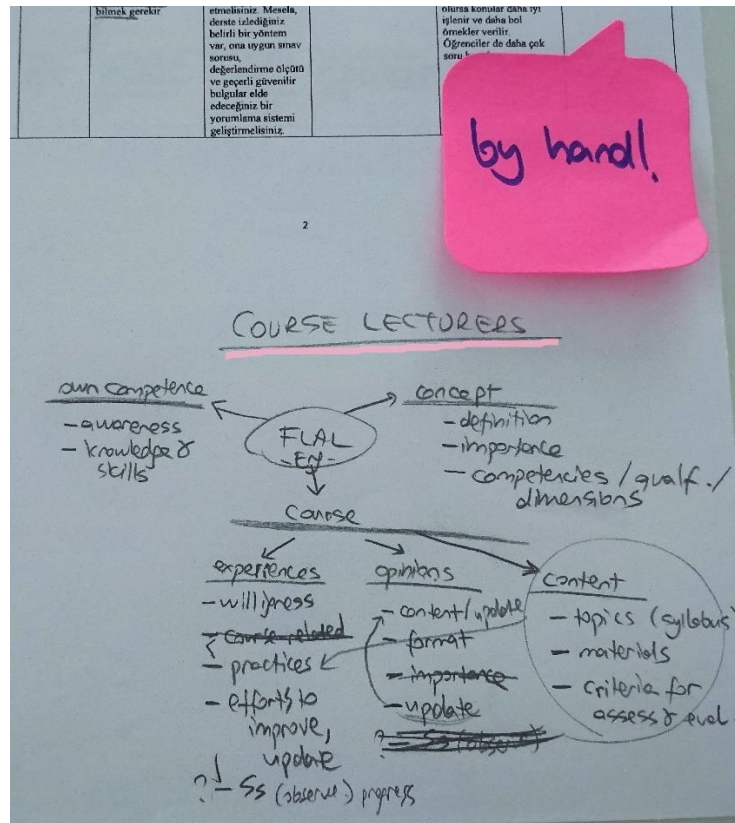
Course Title	Course Code	Semester	Required/Elective	Theory + Practice	ECTS
English Language Testing and Evaluation	İNÖ401	VIII. SEMESTER	Required Courses	3+0	5.0
Language of Instruction	İngilizce				
Course Type	Required Courses				
Course Instructor(s)	YARDIMCI DOÇENT DOKTOR ALİ MERÇ, OKUTMAN NESLİHAN AYDEMİR, DOÇENT DOKTOR ÖZGÜR YILDIRIM, DOÇENT DOKTOR ÖZGÜR YILDIRIM, YARDIMCI DOÇENT DOKTOR GONCA SUBAŞI, YARDIMCI DOÇENT DOKTOR GONCA SUBAŞI, OKUTMAN NESLİHAN AYDEMİR, DOÇENT DOKTOR ÖZGÜR YILDIRIM				
Mode of Delivery	The mode of delivery of this course is face to face.				
Prerequisites	There is no prerequisite or co-requisite for this course.				
Courses Recommended	None				
Reading List					
Assessment Methods	1 Midterm, 1 Final, 1 Assignment				
Work Placement	None				

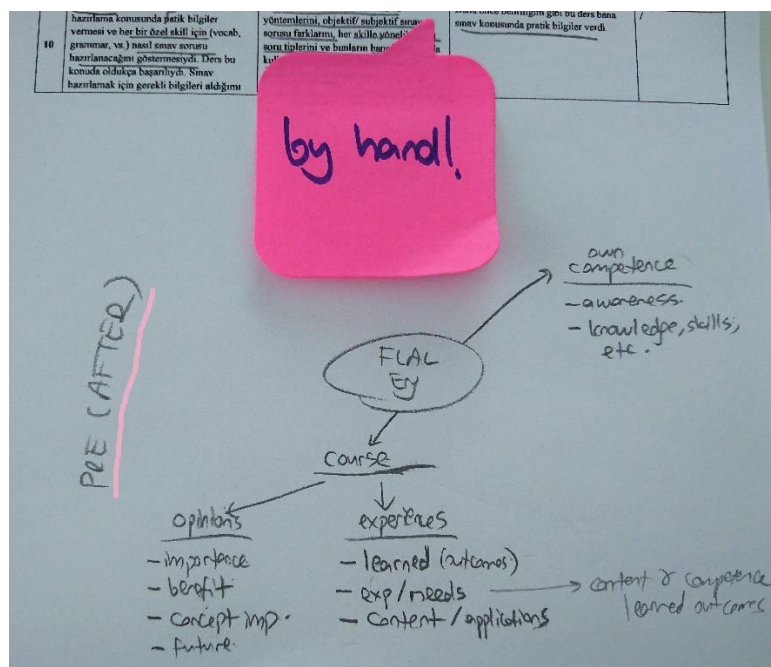
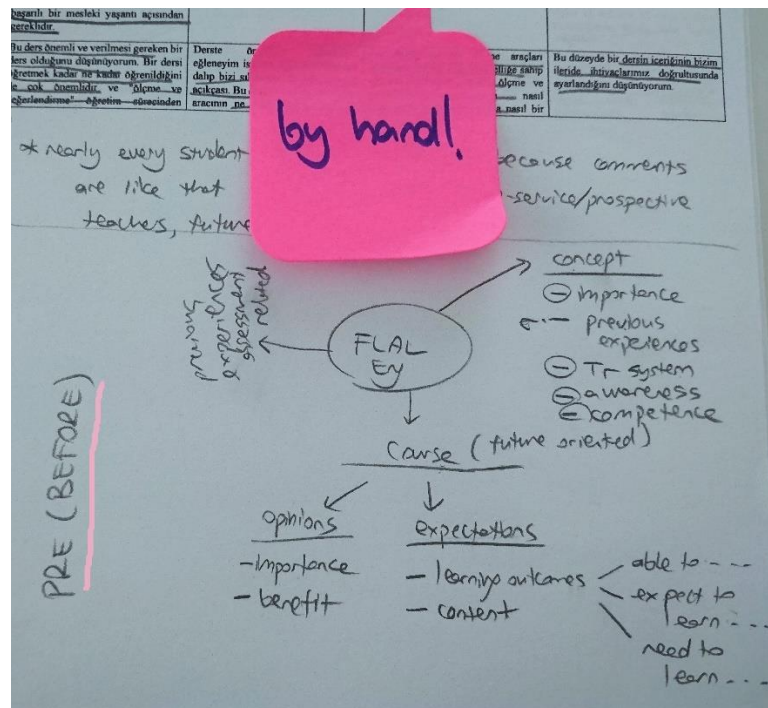
- Description
- Content
- Learning Outcomes
- Learning Activities and Teaching Methods
- Course's Contribution to Prog.
- Assessment Methods

Content

Weeks	Topics
1. Weeks	Introduction of the course: aim and course material
2. Weeks	Introduction to language testing
3. Weeks	Approaches to language testing
4. Weeks	Steps of writing tests. Objective testing
5. Weeks	Criteria and types of tests. Objective testing
6. Weeks	Criteria and types of tests. Objective testing
7. Weeks	Revision
8. Weeks	Revision
9. Weeks	Tests of grammar and usage
10. Weeks	Testing vocabulary
11. Weeks	Listening comprehension tests
12. Weeks	Oral production tests
13. Weeks	Testing reading comprehension
14. Weeks	Testing the writing skills

APPENDIX – 13: The Data Analysis Performed by Hand (The First Impression)





APPENDIX – 14: The Code-Theme Relationship by NVivo (The Last Version)

Name	Files	References
FLAL-Eng (Pre-service BEFORE)		1299
FLAL-ELTE Concept		253
ELTE Course-related Issues		1046
FLAL-Eng (Pre-service AFTER)		1653
FLAL-ELTE Concept		617
ELTE Course-related Issues		1036
FLAL-Eng (Course Lecturers)		295
FLAL-ELTE Concept		93
ELTE Course-related Issues		202
FLAL-Eng (Alumni)		367
FLAL-ELTE Concept		115
Experiences about ELTE (past & present)		252

Name	Files	References
FLAL-Eng (Course Lecturers)		295
ELTE Course-related Issues		202
Experiences about ELTE course		146
Course content & practices		99
Criteria for assessment & evaluation		28
Materials		36
Topics (syllabus)		35
Observations of students' progress		37
Changes after the course		15
Competencies after the course		22
Reasons to teach		10
General opinions about ELTE course		56
General opinions of content		21
Existing content sufficient		4
Own course content - not CoHe		4
Should update content		13
General opinions of its place, length, number, etc.		35
(In)sufficiency of course		22
Semester-year suitability & reasons		13
FLAL-ELTE Concept		93
Definition of ELTE		11
Definition of FLAL		29
Importance of FLAL-ELTE		19
Opinions about qualifications of FLAL-ELTE		14
Perceptions of own FLAL		20
Perceived competency level		10
Teaching experience		10

Nodes			
Name	Files	References	
FLAL-Eng (Alumni)		3	367
Experiences about ELTE (past & present)		2	252
ELTE course-related experiences (previous experi		2	198
General benefits of ELTE course		2	98
Acquired knowledge & skills (content lea		2	37
Preparing for future		2	3
Responding to needs		1	19
Showing own FLAL		2	3
Sufficiency of the course content		2	36
Shortfalls of ELTE course		2	100
Shortcomings		2	12
Suggestions		2	88
Practices of using ELTE course in career life (curr		2	54
Able to use ELTE info in class (how)		2	27
Difficulties in using		2	27
Difficulties as external factors		2	9
Difficulties as internal factors		2	18
FLAL-ELTE Concept		2	115
Definition of ELTE		1	44
Importance of ELTE		1	27
Perceptions of own FLAL		1	44
Perceived competency level		1	22
Teaching experience		1	22

Nodes			
Name	Files	References	
FLAL-Eng (Pre-service BEFORE)		2	1299
ELTE Course-related Issues		2	1046
Expectations & needs of course itself (learning outcomes)		2	492
General opinions of ELTE course (importance & benefits)		2	554
Negative attitude		2	12
No idea		2	4
Positive attitude		2	538
FLAL-ELTE Concept		2	253
Competence-Awareness in general		2	59
Importance of FLAL-ELTE		2	184
Previous Experiences of ELTE		2	10

Nodes			
Name	Files	References	
FLAL-Eng (Pre-service AFTER)		3	1653
ELTE Course-related Issues		3	1036
Experiences about ELTE course		2	449
Acquired knowledge & skills (competencies gained)		2	449
General opinions of ELTE course		3	587
Importance & benefits of ELTE course		2	306
Expectations & needs met		1	137
Reasons for importance & benefit of course		2	169
Opinions of content & course		2	281
Negative attitude		2	58
No idea		1	1
Positive attitude		2	222
FLAL-ELTE Concept		3	617
Importance of FLAL-ELTE		2	263
Perceptions of own FLAL		2	354
Perceived competency level		2	354
Comments of groups about own levels		2	208
Distribution of levels		1	146

APPENDIX – 15: The Findings from Course Lecturers (Detailed Version)

Name	References
FLAL-Eng (Course Lecturers)	295
ELTE Course-related Issues	202
Experiences about ELTE course	146
Course content & practices	99
<i>Criteria for assessment & evaluation</i>	28
<i>Exam</i>	16
Exam type..multiple-choice or open-ended	3
Final exam	6
Midterm exam	6
<i>Other tasks</i>	12
Assignment as language test preparation	6
Exams prepared & administered	1
Presentation	1
Project-based assessment	1
Quiz	1
Weekly reflection	2
<i>Materials</i>	36
<i>Added materials</i>	15
Homework exercises added	1
Not add any material	1
Other books & articles added	4
Reinforcing exercises added	1
Samples & problems of assessment added	1
Samples of standardized tests added	1
Samples of test items & questions added	5
Stats added	1
<i>Book used & reasons</i>	21
Bachman & Palmer's book	1
Brown's book	1
Genesee & Upshur's book	1
Heaton's book	5
Hughes' book	2
<i>Reasons to choose the book</i>	11
According to learning outcomes	1
According to own knowledge	1
According to Ss' needs-levels	2
As a result of resource searching	1
Basic but comprehensive topics	1
Compatible with methods-courses	1
Practical samples	4

Topics (syllabus)	35
Added topics	15
Alternative assessment added	2
CEFR assessment added	1
Dynamic assessment added	1
Evaluation of test results added	1
Integrated language assessment added	1
Internet.computer-based testing added	2
Item analysis added	1
Not add new topic	3
Practices of test preparation added	1
Samples of tests according to age & purposes added	1
Scale development added	1
Existing topics	20
Alternative assessment	1
Approaches to language testing	1
Appropriate interpretation of assessment results	1
Assessing young learners	1
Assessment of language skills & areas	5
Basic knowledge & methods of ELTE	2
CEFR assessment	1
Definition of testing & assessment	1
Internet-based testing	1
Methods of assessment	1
Principles of language testing (reliability, validity, etc.)	1
Selection of appropriate evaluation criteria	1
Test preparation techniques	1
Types of test items & questions	2
Observations of students' progress	37
Changes after the course	15
Being aware of alternative assessment	1
Being interested in assessment topics	1
Being more aware of ELTE	7
Criticizing existing assessment procedures	1
Criticizing previous assessment procedures	1
Gaining general knowledge of ELTE	1
Gaining perspective about evaluation	1
Positive effect on teaching	1
Transformation from methods to assessment	1
Competencies after the course	22
Able to find appropriate assessment criteria	1
Able to make adaptations about testing	1
Able to select appropriate test items	1
Able to write test instructions	1

Feeling more competent	1
Gaining theoretical & practical knowledge	2
Having basic theoretical knowledge, lacking practical skills	3
Having required competencies if attendance done	1
Having theoretical knowledge only	1
Improved in writing test items	1
Increasing practical skills	1
Learning how to prepare & administer tests	2
More experience needed	3
Need more practice	2
Paying attention to test timing	1
Reasons to teach	10
Expert area	2
Program need	2
Willingness-interested-volunteer	6
General opinions about ELTE course	56
General opinions of content	21
<i>Existing content sufficient</i>	4
Evaluating & interpreting results of assessment	1
Measuring language skills & areas	2
<i>Own course content - not CoHE</i>	4
<i>Should update content</i>	13
Assessment of young learners needed	2
CEFR assessment needed	1
General knowledge of ELTE required	1
Learning-oriented assessment required	1
More practical issues should be	1
New topics as alternative assessment techniques should be	2
Performance-based assessment needed	1
Standardized-tests vs classroom-based assessments should be	1
Test preparation techniques according to age & level needed	1
Test preparation techniques for ESP needed	1
General opinions of its place, length, number, etc.	35
<i>(In)sufficiency of course</i>	22
Coordinating with practicum needed	4
Elective courses suggested	1
Insufficient due to not transferring OMB-Tr to ELTE	1
Need to add as a dimension to all methods-courses	1
New course on more practice required	1
One course insufficient & why	4
One course sufficient	3
Sufficient if associated with OMB-Tr course	4
Two following courses needed	3
Semester-year suitability & reasons	13

FLAL-ELTE Concept	93
Definition of ELTE	11
Decision-making about Ss' language progress	1
Evaluating language learning & teaching process	3
Measuring language proficiency (knowledge & skills) according to aims (Whether learning outcomes achieved)	6
To determine Ss' language level & deficiencies	1
Definition of FLAL	29
Ability to evaluate	3
Ability to interpret test results	4
Ability to make decisions according to process	1
Ability to practice assessment procedures	4
Ability to prepare tests	2
Ability to share within code of ethics	1
Ability to use concepts of ELTE	1
Having basic principles & methods of ELTE	4
Having knowledge & skills of ELTE	3
Rationale behind FLA	1
Stats knowledge for evaluation	1
Understanding concepts of ELTE	4
Importance of FLAL-ELTE	19
As a tool of teaching process	2
As important as teaching	1
Assessment embedded in methods-courses	1
Continuous data collection of Ss	1
Decision-making & planning	3
Embedded in all learning & teaching process	1
Feedback about Ss' language deficiencies	1
Inseparable - teaching & assessment	2
Positive effect to teaching	1
Proper implementation	1
To save progress & process notes	1
To see level & deficiencies of Ss & Ts	2
Tool as feedback	2
Opinions about qualifications of FLAL-ELTE	14
Area experience	4
Stats knowledge	1
Subject-content knowledge	9
Perceptions of own FLAL	20
<i>Perceived competency level</i>	10
High	10
<i>Teaching experience</i>	10
0-5 years	3

10-15 years	6
16-20 years	1

APPENDIX – 16: The Findings from Novice EFL Teachers (Detailed Version)

Name	References
FLAL-Eng (Alumni)	367
Experiences about ELTE (past & present)	252
ELTE course-related experiences (previous experiences)	198
General benefits of ELTE course	98
<i>Acquired knowledge & skills (content learned)</i>	37
Able to assess language skills & areas	7
Able to evaluate according to language testing criteria	1
Able to prepare tests for different age & levels	3
Alternative methods learned	1
ELTE criteria learned	6
Helpful in using appropriate language assessment	3
Knowing how to behave against errors	2
Knowledge of preparing tests gained	3
Learning different language testing tools	2
Learning how to evaluate	3
Planning and applying language tests	6
<i>Preparing for future</i>	3
<i>Responding to needs</i>	19
No	2
So-so	1
Yes	16
<i>Showing own FLAL</i>	3
<i>Sufficiency of the course content</i>	36
Adequate for future career	9
Oriented to real life	3
Ss active in the course	1
Sufficient content	22
Useful in language assessment knowledge	1
Shortfalls of ELTE course	100
<i>Shortcomings</i>	12
Acquiring only theoretical knowledge but practical missing	8
Class hours inadequate	2
Imaginary practices in the course	1
Unsuitable as last year course	1
<i>Suggestions</i>	88
Activities as preparing & evaluating tests needed	6
Coordination with practicum needed	4
Different perspectives for different schools needed	1
Knowledge of test format needed	1
More evaluation & interpretation tasks needed	2
More practical knowledge needed	5

More practices on sample tests needed	7
More practices on testing different levels	1
More productive skills testing training needed	3
More realistic practices needed	1
Other language assessment methods needed	1
Performance-based testing needed	1
Preparing rubrics needed	2
Simplifying content	2
Simulations can be added	5
Two following courses needed	1
<i>In-service training</i>	45
<i>No need-content sufficient</i>	1
<i>Positive approach</i>	18
<i>Reasons for positive approach</i>	26
Being incompetent in ELTE	2
Different implementations in university training	1
For those who did not take ELTE training	2
To be able to adapt learned to real life	1
To create awareness	1
To develop teaching techniques	1
To evaluate learning outcomes better	2
To fix wrong applications	5
To know ELTE better	2
To know how to test language proficiency better	2
To maximize experiences	2
To remind the course content-knowledge of ELTE	4
To update ELTE knowledge & skills	1
Practices of using ELTE course in career life (current experiences)	54
Able to use ELTE info in class (how)	27
Appropriate use of language assessment	7
Helpful in getting feedback	1
Increase Ss' awareness of language	1
Paying attention to language testing criteria	7
Trying to use learned knowledge & skills	10
Using for distribution of test items	1
Difficulties in using	27
<i>Difficulties as external factors</i>	9
Local conditions	4
Tr testing system	1
Turkish context-EFL	2
<i>Difficulties as internal factors</i>	18
Forgetting course content	3
Hard to practice	8

Need more experience	2
Unable to adapt to real life	5
FLAL-ELTE Concept	115
Definition of ELTE	44
Analysis of test results	3
Interpretation of language proficiency	1
Knowledge of ELTE itself	1
Oral or written language exams	2
Planning learning outcomes for language	1
Preparation & administration of tests	1
Reliable & valid testing tools	2
Reporting evaluation findings	1
To determine language proficiency	8
To determine Ss' language deficiencies	1
To determine whether training useful	1
To evaluate Ss' language proficiency	3
To grade knowledge in numbers	1
To measure language skills & areas	5
To measure Ss' language progress	1
To measure the use of language for expression	7
Understanding whether aims achieved (feedback)	5
Importance of ELTE	27
Being multi-dimensional	2
ELTE useful if informal	1
Helping to determine language progress	2
Shaping education	1
Showing whether aims achieved	5
To check language proficiency	6
To cover lacking parts	4
To evaluate language progress	2
To get feedback about language progress	3
To interpret the results	1
Perceptions of own FLAL	44
<i>Perceived competency level</i>	22
High	11
Low	2
Moderate	9
<i>Teaching experience</i>	22
1 year	3
2 years	7
3 years	5
4 years	4
5 years	3

APPENDIX – 17: The Findings from Pre-service ELT Teachers (Before the Course)
(Detailed Version)

Name	References
FLAL-Eng (Pre-service BEFORE)	1299
ELTE Course-related Issues	1046
Expectations & needs of course itself (learning outcomes)	492
No expect or need	3
Rich content expected	3
To be able to evaluate fairly	5
To be able to grade properly	5
To be able to practice assessment	21
To be competent in assessing language proficiency	5
To be competent in assessment tools	3
To be competent in testing & evaluation	29
To enjoy the course	6
To gain awareness of assessment concept	22
To gain knowledge & ability of assessment (prepare, design, evaluate, etc.)	64
To get high score to pass	6
To improve assessment-related knowledge & skills	5
To learn & be able to how to assess effectively	78
To learn assessment criteria	37
To learn assessment techniques	57
To learn background of assessment	1
To learn English-related assessment	39
To learn future-related knowledge & skills	23
To learn how to evaluate (own teaching)	5
To learn how to evaluate (SS, process, etc.)	38
To learn how to give feedback via assessment	5
To learn other purposes of assessment	1
To learn practical side	9
To learn stats	1
To learn the relationship between teaching & assessment	8
To learn washback effect	5
To need good samples of tests	1
To remember methods courses & summary of training	7
General opinions of ELTE course (importance & benefits)	554
<i>No idea</i>	4
<i>Negative attitude</i>	12
Course benefit questionable	1
Course not important	4
Not practical	1
Only as a tool of reminding	2

Similar to OMB-Tr course	2
Too late to take the course	2
Positive attitude	538
Reasons for importance & benefit of course	226
Assessment-testing compulsory	13
Beneficial to reliable & valid test design	6
Change the perception of assessment	3
Complementary to methods-courses	18
Different from OMB-Tr course	6
Filling the gaps of teaching competence	9
Know nothing about how to test-assess	9
Learning English-related assessment	9
Taking course from an experienced teacher	11
Testing system concerned	9
To be aware of assessment concept	10
To improve ourselves	7
To learn how to test properly (theory+practice)	42
Useful for future teaching career	74
Yes, course beneficial	159
Yes, course important	153
FLAL-ELTE Concept	253
Competence-Awareness in general	59
Ability to prepare tests according to criteria	4
Ability to teach as well as assess	4
Need to be competent in how to test-assess (knowledge)	38
Need to know how to evaluate	4
Need to know testing techniques	5
To test & evaluate in a fair way	4
Importance of FLAL-ELTE	184
Assessment as important as teaching	36
Classroom vs standardized testing	1
Need to have the knowledge	12
Summative vs formative assessment	2
To assess effectiveness of teaching	36
To assess whether Ss learn or not	39
To determine Ss' level	11
To distribute scores in a fair way	1
To evaluate test items prepared	1
To find out Ss deficiencies	11
To find out whether aims achieved	2
To give feedback	6
To guide teaching as a result of assessment	7
To make Ss more successful	6

To show Ss' mistakes	3
To show Ss' progress	8
To test the quality of assessment	2
Previous Experiences of ELTE	10
Complaining about evaluation	1
Different kinds of wrong exams	6
Previous teachers' lack of ELTE competence	1
Taking too many exams so far	2

APPENDIX – 18: The Findings from Pre-service ELT Teachers (After the Course)
(Detailed Version)

Name	References
FLAL-Eng (Pre-service AFTER)	1653
ELTE Course-related Issues	1036
Experiences about ELTE course	449
Acquired knowledge & skills (competencies gained)	449
Approaches to testing learned	2
Assessment criteria & principles learned	93
Both knowledge & skills gained	9
Different kinds of testing methods learned	41
Gained knowledge of testing each skill	84
How to assess properly learned	33
How to design a test learned	69
How to evaluate learned	34
How to motivate via testing learned	2
Purposes of tests learned	6
Test types & items learned	34
Tests for different levels learned	14
The connection between testing & teaching learned	4
Views of testing changed	24
General opinions of ELTE course	587
Importance & benefits of ELTE course	306
<i>Expectations & needs met</i>	137
No	2
So-so	7
Yes	128
<i>Reasons for importance & benefit of course</i>	169
Basic knowledge & skills given	24
Complementary to methods courses	6
Different from OMB-Tr course	8
Engaging in fruitful lessons	6
Learned a lot & created awareness	51
Transferring knowledge between courses(OMB-ELTE)	2
Useful for future career	72
Opinions of content & course	281
<i>Negative attitude</i>	58
<i>Content shortcomings & course shortfalls</i>	57
Last semester too late	6
No info about test format, etc.	1
Not updated content	8
Only one assignment insufficient	2
Only testing language skills	1

Only theory, no practice	10
Presentations done by students	1
<i>Suggestions</i>	28
More checklists needed	1
More class hours needed	1
More examples for integrated skills	1
More samples on good-bad items of tests	2
More samples on table of specifications needed	2
Need more exercises to practice	13
Other topics about ELTE needed	1
Should be integrated into practicum	1
Should be updated	6
<i>No, insufficient content</i>	1
<i>No idea</i>	1
<i>Positive attitude</i>	222
<i>Opinions of content sufficiency</i>	91
Both theory & practice	6
Course instructor good	20
Coursepack comprehensive	11
Examples for how to evaluate	1
Examples for testing each skill	19
Good & bad sides of assessment	1
Suitable teaching of topics	7
Timing good	1
Topics according to levels	1
Topics comprehensive	24
<i>Yes, sufficient content</i>	131
FLAL-ELTE Concept	617
Importance of FLAL-ELTE	263
Assessment as important as teaching	32
Backwash effect	6
Need to acquire the specific competence	38
Needed in not only English but also other fields	10
Scoring not essential	3
Serving as feedback	14
To assess in a fair way	7
To assess language proficiency	17
To assess Ss' success	9
To evaluate effectiveness of teaching	22
To find needs of Ss	2
To find out language deficiencies	9
To guide teaching	18
To improve Ss more	18

To see whether Ss learn or not	51
To update oneself	7
Perceptions of own FLAL	354
<i>Perceived competency level</i>	354
<i>Comments of groups about own levels</i>	208
<i>High-comments</i>	41
H-Ability to assess properly	12
H-Attend lessons regularly	3
H-Being active in lessons	1
H-Course good	5
H-Feel competent	19
H-Improve more with experience	1
<i>Low-comments</i>	7
L-At the phase of only understanding	1
L-Feel incompetent	2
L-Improve better after practice	2
L-No experience	2
<i>Moderate-comments</i>	160
M-Course good	14
M-Develop more with career	35
M-Difficulties in reliability & validity	1
M-Difficulties in scoring	1
M-Lose attention due to anxiety of scoring	1
M-Lose knowledge due to not repeating	1
M-Need practice	33
M-Need to be expert	13
M-Need to improve about productive skills	1
M-No experience	18
M-Not attend classes regularly	5
M-Only one course inadequate	2
M-Theory good but practice missing	35
<i>Distribution of levels</i>	146
High	33
Low	4
Moderate	109

APPENDIX – 19: The Findings of the Document Analysis

Categories							
1. Name of University	<i>Anadolu</i>	<i>Bogazici</i>	<i>Middle East Technical</i>	<i>Hacettepe</i>	<i>Eskisehir Osmangazi</i>	<i>Marmara</i>	<i>Bursa Uludag</i>
2. Status of University	State	State	State	State	State	State	State
3. Title of the Course	English Language Testing and Evaluation	Language Assessment	English Language Testing and Evaluation	Foreign Language Measurement and Evaluation	English Language Testing and Evaluation	Assessment and Evaluation in Foreign Language Teaching	English Language Testing and Evaluation
4. Type of the Course	Compulsory /Required	Compulsory	Compulsory	Compulsory/Must	Compulsory	Compulsory	Compulsory
5. Language of Instruction	English	English	English	English	English	English	English
6. Semester (Course Given)	8 th semester (4 th year-spring)	7 th semester (4 th year-fall)	7 th semester (4 th year-fall)	8 th semester (4 th year-spring)	8 th semester (4 th year-spring)	8 th semester (4 th year-spring)	8 th semester (4 th year-spring)
7. Hours/ECTS of the Course	3/5	4/6	3/4.5	3/5	3/3	3/5	3/5
8. Mode of Delivery / Learning Activities & Teaching Methods	Face-to-Face N/A	N/A	N/A	Face-to-Face Lecture, discussion and question & answer	N/A	Classroom presentation, group work, practical application, class discussion	Face-to-Face N/A

9. Instructor(s) of the Course	4 instructors	4 instructors	2 instructors	1 instructor	1 instructor	1 instructor	2 instructors
10. Book(s) Read in the Course	Heaton, J. B. (2011). Writing English Language Test.	N/A	N/A	Heaton, J. B. (1990). Writing English Language Test.	1) Bachman, L. (1990). Fundamental considerations in language testing. 2) Alan, D. (1990). Principles of language testing. 3) Harris, D. P. (1969). Testing English as a second language. 4) Hughes, A. (1988). Testing English for university study. 5) Heaton, J. B. (1990). Classroom testing. 6) Heaton, J. B. (1990). Writing English Language Test.	1) Brown, H.D. (2003). Language assessment: principles and classroom practice. 2) Hughes, A. (1989). Testing for language teachers. 3) Genesse, F., & Upshure, J. A. (1996). Classroom- based evaluation in SLA.	N/A

7) Cyril, W. & Roberts, J. (1994). Assessment in ELT.							
11. Content of the Course	The same as its topic list	N/A	Types of tests; test preparation techniques for skills of English; practice of preparing various types of questions; evaluation and analysis techniques; statistical calculations	N/A	The same as learning outcomes	The same as its topic list	N/A
12. Topic List of the Course (Outline)	Introduction ; Approaches to language testing; Objective testing; Criteria (reliability, validity, etc.) and types of	Theories of measurement, assessment and evaluation; construction and evaluation of tests for assessing second	N/A	Introduction; Approaches to language testing; objective testing; tests of grammar, vocabulary, listening, speaking, reading, writing;	Basic concepts of language testing; types of tests; major aspects of English language testing; procedures in preparing different test items; practice	Teaching & testing; kinds of tests and test tasks; stages of test construction; testing grammar, vocabulary, reading, listening, writing, oral	N/A

	tests; tests of grammar, vocabulary, listening, speaking, reading and writing	language skills; reliability, validity and backwash; statistical analysis of test scores and interpretation of results		criteria (reliability, validity, etc.) and types of tests	of preparing questions; test preparation techniques for language skills; testing writing, reading, speaking, listening; analyzing techniques and statistical calculations	ability; interpreting test scores; standardized tests	
13. Assessment Criteria of the Course	Term: 60% Final: 40% ----- ---- Midterm exam: 30% Homework: 30% Final exam: 40%	N/A	N/A	Term: 60% Final: 40% ----- -- Attendance: 10% Assignments: 40% Midterm: 10% Final exam: 40%	Term: 1 midterm exam & 2 homework Final: 1 final exam	Term: 40% Final: 60% ----- Term: Midterm: 80% Homework: 20% ----- Final: Exam: 60% Project: 40%	Term: 40% Final: 60%
14. Objective(s) of the Course	N/A	N/A	- learn basic terms and concepts of language testing, - engage in practices of	- develop learners' concept of testing and making him aware of	same as learning outcomes	- to increase students' awareness, - to enable students to	N/A

			language assessment, - perform statistical analysis, - design, evaluate and implement language testing instruments	different types of, producing them when necessary		design, implement and evaluate a variety of testing instruments	
15. Learning Outcomes of the Course (Basic Key Words)	- basic principles of language testing, - approaches to language testing, - analyze and prepare appropriate language tests, - interpret test scores and give feedback	N/A	- basic terms and concepts of language testing, - knowledge, selection and design of language tests, - write, implement, adapt and evaluate language tests, - perform statistical analysis of test data, - advantages and disadvantages of using published and class teacher-written tests	- approaches to language testing, - objective testing, - different types of language tests, - prepare various language tests	- design language tests appropriately, - prepare tests of language skills, and grammar and vocabulary, - criteria of testing, - basic statistics	- be able to report interpretation, - design and implement tests, - evaluate advantages and disadvantages of test techniques, - compare test types and methods, - define and identify testing methods	N/A

APPENDIX – 20: The Original Turkish Versions of the Quotes

The Quotations of Course Lecturers

O.CL.6: “Yabancı dil öğrenmekte olan kişinin amaçlara göre dil yeterliliğini bilgi ve beceri kullanımını bazında değerlendirmektir.”

‘Testing and evaluation is assessing language proficiency of a foreign language learner on the basis of its use with respect to learning purposes.’ (O.CL.6)

E.CL.8: “Yabancı dil yeterliliğini güvenilir ve geçerli araçlarla süreç ve son ölçüm olarak etkili değerlendirme sürecidir.”

‘Testing and evaluation is assessing effectively language proficiency as carrying out both summative and formative assessment by means of reliable and valid testing tools.’ (E.CL.8)

A.CL.4: “Yabancı dilde öğretilen dil alan ve becerilerinin öğrenilip öğrenilmediğine bakmak, öğretim sürecinin öğretmen, öğrenci ve sürecin kendisi açısından nasıl gittiğine bakmaktır.”

‘ELTE can defined as assessing whether language skills and areas in a foreign language are learned or not, and evaluating the teaching process from the perspective of teachers themselves, students and process itself.’ (A.CL.4)

A.CL.3: “Bu kavramın alt başlıkları da var aslında. Öncelikle yabancı dilin nasıl ölçülebileceği konusunda bilgi sahibi olmalısınız. Daha sonrasında uygun bir sınav hazırlama becerisine ve değerlendirirken kullanacağınız istatistiki bilgiyi doğru yorumlama becerisine sahip olmalısınız. Ayrıca bu sınavı hazırlarken izleyeceğiniz birtakım kuralların ve prensiplerin olmasına dikkat etmelisiniz. Mesela, derste izlediğiniz belirli bir yöntem var, ona uygun sınav sorusu, değerlendirme ölçütü ve geçerli güvenilir bulgular elde edeceğiniz bir yorumlama sistemi geliştirmelisiniz.”

‘The concept of FLAL has actually subcategories. First, you need to have the knowledge about how to test language proficiency. Then, you need to have the ability to design language tests and the competence to interpret assessment results appropriately in the light of statistical knowledge. Besides, you need to pay attention certain principles and rules while preparing such tests. For example, there is a certain way of teaching style used in your course, and accordingly, you have to develop an evaluation system in which you can prepare related questions, construct evaluation criteria and obtain reliable and valid findings.’ (A.CL.3)

B.CL.5: “Okuryazarlık, bir mesleğin ya da uygulamanın gereklerini yerine getirecek kadar iyi bilmek anlamına da gelebiliyor. Ben dil bağlamında düşünüp, çeşitli amaçlar için çeşitli bağlamlarda yabancı dil konuşan bireylerin düzeylerini amaca uygun bir biçimde değerlendirme uzmanlığına sahip olmak olarak tarif etmek istiyorum.”

‘Literacy can refer to the expertise in any profession or application which needs to be knowledgeable about its requirements to be able to fulfill them. In terms of language concept, I can define literacy as having the expertise in an appropriate assessment procedure through which foreign language learners’ proficiency is assessed with respect to the aims.’ (B.CL.5)

U.CL.10: “Yabancı Dilde Ölçme-Değerlendirme bilgisin ne zaman ve neden ihtiyaç duyulduğunu, doğru Ölçme-Değerlendirmenin nasıl yapılacağını ve sonuçların nasıl

yorumlanacağı ve saklanabileceğini, nasıl değerlendirileceğini ve etik bir biçimde nasıl paylaşılacağını bilmektir.”

‘FLAL is to know when and why assessment knowledge is needed, how to perform effective testing and evaluation, how to interpret and store the results, and how to share the results ethically.’ (U.CL.10)

A.CL.3: “Çok önemli. Öğretme ve değerlendirme ayrılmaz bir bütün aslında. Uygun değerlendirme yapmadan öğretmede yeni adımlar atamayız, etkili planlar yapamayız. Öğrencilerin İngilizce alanındaki temel eksiklikleri ile geribildirim bize en etkin, çabuk ve net olarak uyguladığımız sınavlar ve ödevler verir.”

‘Assessment is very important. Testing-evaluation is an inseparable concept. If we did not assess properly, we would not be able to take new steps in teaching and to prepare effective plans. The feedback about the main deficiencies of students in English is provided most efficiently and quickly by tests and assignments.’ (A.CL.3)

H.CL.7: “Diğer alanlarda olduğu gibi öğrenci, öğretmen ve program hakkında dönüt vermek.”

‘As in other fields [apart from ELT], the purpose of assessment is to give feedback about students, teachers and program itself.’ (H.CL.7)

M.CL.9: “Evet gerektiriyor. Hem teorik hem de pratik anlamında teknik bir konu.”

‘Foreign language assessment needs expertise in this field. It is a technical subject in terms of both theoretically and practically.’ (M.CL.9)

B.CL.5: “Tabii ki. Temel kavram ve prensiplerin yanı sıra, özellikle istatistik yöntemlerin de bilinmesi gerekiyor. Kitap açıp okuyarak öğrenilecek bir branş değil.”

‘Foreign language assessment requires a specific qualification in this area. In addition to the basic concepts and principles of testing, it is important to know especially statistics. It is not a branch to be learned by reading a book.’ (B.CL.5)

E.CL.8: “Genel hatlarıyla; What is Testing? What is assessment? Principles of Classroom Testing (Reliability, Validity etc), Standardized Testing, Testing Grammar/Vocabulary, Assessing Listening, Assessing Reading, Assessing Writing, Assessing Speaking, Integrated Testing and Assessment”

‘My syllabus includes the topics of what is testing, what is assessment, principles of classroom testing (reliability, validity, etc.), standardized testing, testing grammar/vocabulary, assessing listening, assessing reading, assessing writing, assessing speaking, integrated testing and assessment.’ (E.CL.8)

E.CL.8: “Öğrencilere uygulama olarak test hazırlama ve alternatif değerlendirme ekledim. Ayrıca, integrated testing üzerine daha çok duruyorum çünkü dilin bütüncül yapısını ve dil yeterliliğini belirlemede bütünleşik değerlendirmenin önemini öğrencilerin de anlamasını istiyorum.”

‘I added the topics of preparing language tests practically and alternative assessment. Besides, I focus mostly on integrated testing because I want students to understand the importance of integrated testing while assessing the overall structure of language itself and language proficiency.’ (E.CL.8)

A.CL.3: “Çeşitli yaş grupları ve amaçlar için hazırlanmış sınav örneklerini ekledik çünkü ihtiyaç olursa öğrencilerin ellerinde güzel bir örnek olsun istedik.”

‘We added sample tests for various age groups and purposes because we wanted students to have a good coursepack if they needed.’ (A.CL.3)

A.CL.2: “Pratik uygulama örnekleri içeren, sadece teorik bilgi vermeye bir kitap olması sebebiyle seçildi.”

‘[We chose this book [Heaton’s book] as the coursebook because] it consists of practical examples, and thus, it does not only provide theoretical information.’ (A.CL.2)

A.CL.1: “Teorik bir ara sınav ve ilgili dil alan ve becerilerini ölçen testlerin öğrenciler tarafından hazırlanması”

‘[As for the assessment and evaluation criteria of ELTE course], students must enter a theoretical test as the midterm exam and prepare a test that measures language skills as the final exam.’ (A.CL.1)

O.CL.6: “Kesinlikle. Bir farkındalık oluyor ve sınav hazırlama ve sınav yapma konusunda aydınlanıyorlar. Aldıkları her sınavı o gözle eleştiriyor, hazırladıkları sınava önem veriyorlar.”

‘Definitely, there is a change at the end of the course. Students become more aware of assessment and are enlightened about exam preparation and administration. They criticize each exam they take and give importance to each exam they prepare.’ (O.CL.6)

B.CL.5: “Kabaca gözlemim, dördüncü sınıf dersi olarak verilen bu derste, öğrencilerin o zamana kadar oluşturduğu öğretim yöntemlerinin değerlendirme yöntemleri şekline verilmesini izliyorsunuz. Çoğu ikisinin arasındaki farkı anladığında bayağı şaşıyor. Bir de bu işin aslında ne kadar da titizlikle yapılması gereken bir iş olduğunu gördüğünde.”

‘My general observation about this course given at the fourth year of teacher training is that students’ teaching methods that they have learned evolve into assessment methods. [...] They are also surprised by the fact that assessment is actually a meticulous job.’ (B.CL.5)

A.CL.3: “Evet, özellikler ödevlerini hazırlarken daha uygun soruları seçebiliyor, gerekli adaptasyonları yapabiliyorlar. Yönergeleri daha düzgün yazıyorlar, zaman ayarlamasına dikkat ediyorlar. Ayrıca uygun değerlendirme kriterini buluyor ya da hazırlarken belli noktalara dikkat ediyorlar.”

‘[After taking this course], students are able to choose more appropriate test items while doing their assignments, and also, they can make necessary adaptations. They are able to write test instructions more properly and pay attention to time allocations among test tasks. Moreover, they can find appropriate evaluation criteria or take into account certain points while preparing such criteria.’ (A.CL.3)

H.CL.7: “Tecrübe eksikliği var ama ders sonunda teorik bilgi oldukça iyi, tecrübeye hazır oluyorlar.”

‘There is a lack of experience [about language assessment], but at the end of the course, students’ theoretical knowledge is quite good and are more ready to assess.’ (H.CL.7)

A.CL.2: “Ölçme Değerlendirme konusuna ilgilim sebebiyle dersi vermek istedim.”

‘Because of my interest in testing and evaluation, I wanted to teach this course.’ (A.CL.2)

A.CL.1: “YÖK'ün önerisinden haberim yok. Kendimiz belirledik. Uygun kitabı seçtik. Okutuyoruz.”

‘I have not heard about CoHE’s program [ELTE course content]. We ourselves [course lecturers] determined the topics. We chose appropriate coursebook and are studying it.’ (A.CL.1)

A.CL.4: “Süreç odaklı alternatif değerlendirme konuları (use of observation checklists, learning journals and logs, portfolios, progress reports and etc) daha fazla yer alabilir.”

‘The topics of process-oriented alternative assessment (use of observation checklists, learning journals and logs, portfolios, progress reports, etc.) should be more emphasized in this course.’ (A.CL.4)

A.CL.1: “Yeterli. Bir de 3. sınıfta Türkçe aldıkları bir ders var. İkisi iyi bir şekilde kotarılsa yeterli olacaktır.”

‘This course [ELTE] is sufficient. There is also a course that is taken in Turkish in the third year of the teacher training program. If these two courses (Turkish general one and ELTE) are learned thoroughly, it will be sufficient in terms of language assessment.’ (A.CL.1)

A.CL.3: “Kendi işledikleri bir ders için staj öğrencilerine gerçek bir sınav hazırlayıp, değerlendirip, sonuçlarını yorumlayabilseler çok daha iyi olurdu.”

‘[Apart from the course], it would be much better in the practicum if students prepared a real English exam, assessed their students and evaluated the results in the lessons they taught.’ (A.CL.3)

A.CL.4: “Aslında bence tek ders yeterli olamıyor, bir dönem daha öğrenilen teoriyi pratiğe dökmeye dayalı, daha çok alternative assessment odaklı bir ders konabilir. Öğrenciler bu dersi genel ölçme değerlendirme şeklinde ana dillerinde de alıyor ama anlamlı bir öğrenme gerçekleştirilemiyor, orda öğrenilenleri bu derse aktaramıyorlar.”

‘In fact, only one course [about ELTE] is not enough; there may be another course which is based on putting knowledge into practice and is more alternative assessment oriented in another semester. Students take a general testing and evaluation course in their mother tongue [Turkish], but they cannot achieve meaningful learning and are not able to transfer what they learn in that course to this ELTE course.’ (A.CL.4)

B.CL.5: “Değerlendirme boyutu aslında bütün metodoloji ve uygulama derslerinin içinde bulunuyor. Her ne kadar öğretme yönelimli (oriented) görev (task) ile değerlendirme yönelimli görevler farklılaşabilse de bahsettiğim derslerin içinde de değerlendirme boyutu ele alınıyor, bilinçli ya da bilinçsiz. Bence dilde ölçme ve değerlendirme temel kavram ve özellikleri öğrenildikten sonra metodoloji derslerine de bir boyut olarak eklenebilir. Tek bir ders her zaman çok az maalesef.”

‘The dimension of assessment is actually embedded in all teaching methods courses. Although learning-oriented and task-oriented assessment can be distinct, the element of assessment is given consciously or unconsciously in such courses [teaching methodology]. In my opinion, after fundamental concepts and principles of testing are learned, assessment might be added as a dimension to methods courses. Only one course for ELTE is inadequate all the time.’ (B.CL.5)

E.CL.8: “Şu anda bulunduğu 4. sınıf 2. dönem uygulama yapma ve farkındalık olarak gayet uygun görünüyor.”

‘The second term of the fourth year [8th semester] seems quite good due to the suitability of performing practical applications as well as creating awareness.’ (E.CL.8)

B.CL.5: “Bence metodoloji derslerinin bir boyutu olarak ele alınırsa 3. sınıftan başlayabilir.”

‘I think if language assessment is taken as a dimension of teaching methods courses, ELTE education may start in the third year.’ (B.CL.5)

The Quotations of Novice EFL Teachers (Alumni)

Low.T.10: “Öğrencinin dildeki yeterlilik düzeyini ölçüp, başarısını değerlendirmektir.”

‘Assessment is to measure the proficiency level of language learners as well as evaluating their successes and progress in that foreign language.’ (Low.T.10)

High.T.8: “Yabancı dilde ölçme-değerlendirme sadece teste veya yazıya bağlı kalmayıp, öğrencinin dili etkin bir şekilde kullanabilme becerilerinin de ölçülmesidir.”

‘Testing and evaluation in a foreign language is not only bound to tests or written exams, but also it means to assess students’ ability to use the target language effectively.’ (High.T.8)

High.T.11: “Dilsel yeteneklere ait kazanımların planlanması ve davranışlarına, kazanımlara ne ölçüde ulaşıldığının test vb ölçme araçlarıyla ölçülüp değerlendirilmesi sürecidir.”

‘Testing-evaluation is the process of planning learning outcomes related to language skills, and to measure and evaluate to what extent these outcomes are achieved by means of testing tools.’ (High.T.11)

High.T.17: “Dil becerilerinin sınanması ile ilgili kavram ve yöntemler hakkında bilgi sahibi olmak, dil yeterliliği ve gelişimi ile ilgili sınav hazırlama süreçlerinin incelenmesi, belirli bir öğrenci grubuna yönelik bir dil testinin hazırlanması- planlama, uygulama ve sonuçların madde analizi ile değerlendirilmesi. Değerlendirme sonuçlarının yazılı bir rapor haline getirilmesi ve sözlü olarak sunulması.”

‘Language assessment means to have knowledge about the concepts and methods related to the testing of language skills, to examine test preparation processes related to language proficiency and development, to prepare a language test for a specific group of students, and to plan, implement and evaluate the results via item analysis. And also, it means to report assessment results in written format and present them verbally.’ (High.T.17)

High.T.6: “Ölçme ve değerlendirme dil öğretiminde kritik bir öneme sahiptir. Öğretme-öğrenme sürecinde dil yeterliliğini sınanması ve geliştirilebilmesi için dönüt verilmesi gerekir. Bu dönütler eksiklerin kapatılmasını sağlayacaktır.”

‘In the teaching/learning context, feedback should be given in order to evaluate and improve language proficiency of learners. So, feedback comes from the assessment results of learners, and is used again to develop their language proficiency. In this way, the shortfalls of the education may be fulfilled.’ (High.T.6)

Moderate.T.18: “Belirlenen kazanımlara ulaşıldıysa yeni kazanımlarla eğitime devam etmek, ulaşamadıysa geriye dönüp ulaşabilmek için çalışmalar, tekrarlar yapmak açısından önemli bir rol oynar.”

‘Assessment plays an important role in terms of carrying on the instruction with new objectives if planned learning outcomes are achieved or going back to the study and making revisions if the objectives are not achieved.’ (Moderate.T.18)

Moderate.T.8: “Ölçme- değerlendirme araçları İngiliz Dili Eğitimi'nde çok yönlü olmalıdır. Çünkü dil öğretimi tek boyutlu düşünülemez. Bir öğrencinin dil gelişimini ölçme işlemleri kelime bilgisi, cümle bilgisi, dinlediğini anlama, okuduğunu anlama gibi çok boyutlu olmalıdır.”

‘Assessment tools should be multi-faceted in ELT because language teaching cannot be thought as being only one dimension. The assessment procedures should be multi-dimensional including a student’s language progress with respect to his/her knowledge of grammar and vocabulary as well as the skills of listening and reading comprehension and the like.’ (Moderate.T.8)

Moderate.T.7: “Elbette işime yaradı. Basitçe söylemem gerekirse sınav nasıl hazırlanır, kapsamı nasıl olmalıdır, sorular nasıl hazırlanmalıdır ve değerlendirme nasıl yapılmalıdır konularında ve bu işlemlerin ayrıntılarında bana yardımcı oldu ve olmaktadır.”

“Surely, this course [ELTE in teacher training] has helped me a lot. Simply put, it has been helpful with respect to such topics as ‘how a language test should be designed, what it should be covered, how the questions/tasks should be organized, and how the interpretation/evaluation should be made’”. (Moderate.T.7)

High.T.4: “İlk deneyimimden yola çıkarak söyleyebilirim ki bu konuda hiç zorlanmadım. Kendi ölçme değerlendirme ölçütlerimi oluşturabilirim. Okuldaki diğer öğretmenlerimiz hazır uygulamaları tercih ederken ben kendi ölçme değerlendirme araçlarımı oluşturdum. Öğrencilerime ve kazanımlarıma uygun olduğu için daha çok başarı elde ettim.”

‘Considering my first experience [about language assessment], I can say that I have not had any difficulty in testing language [at my school]. I am able to construct my own assessment criteria. Although the other teachers in the school [colleagues] prefer ready-made assessment tools, I created my own tests. Since my tests are more appropriate to my students as well as the learning outcomes, I have achieved more success.’ (High.T.4)

High.T.4: “İçeriği gerçek hayata yönelik uygulamalar olduğu için ve öğrencilerin aktif olduğu bir ders olduğu için kazanımlarının daha büyük ve yararlı olduğunu düşünüyorum.”

‘The content of the course [ELTE course] consisted of real-life assessment implementations and the undergraduates were very active in that course. Therefore, I think the course is very beneficial and advantageous. In general, the course was sufficient.’ (High.T.4)

Moderate.T.15: “Yeterlilik düzeyi öğretmenliğe adım atacak öğrenciler için uygundur.”

‘The level of sufficiency [of ELTE course] is appropriate to the students who will be teachers after graduation.’ (Moderate.T.5)

Moderate.T.2: “Karşıladığını düşünüyorum. Dil eğitimindeki beceri alanlarını ölçme ve değerlendirme için gerekli olan kriterleri sağlamıştır.”

‘I think the course [ELTE course] was adequate. It provided us with necessary knowledge and skills about assessing language skills.’ (Moderate.T.2)

Moderate.T.20: “Çok karşıladığını söyleyemem, ölçme ve değerlendirme türleri hakkında bilgi edinip uygulamada eksik kaldığımı düşünüyorum.”

‘I cannot say that this course [ELTE course] did meet my needs and expectations. I only gained the information about assessment evaluation, but I think the course was deficient in terms of practice.’ (Moderate.T.20)

Moderate.T.20: “İçeriği bilgi anlamında doluydu ama pratik kısmı yeterli değildi.”

‘The content of the course [ELTE course] was satisfying in terms of theory but the practical side was not enough.’ (Moderate.T.20)

High.T.1: “Daha çok uygulama olmalı. Mesela her hafta her skill'e özel sınav hazırlanmalı ve değerlendirme kriteri yorumlanmalı.”

‘The [ELTE] course was very theoretical. More practice would have been better. For example, for each language skill, students should prepare an assessment tool every week, and the criteria for the evaluation should be interpreted.’ (High.T.1)

Moderate.T.2: “.....çeşitli tiplerde soru hazırlama alıştırmaları; değerlendirme ve analiz teknikleri; istatistik hesaplamalar.”

‘[To the ELTE course content], exercises of preparing various test items, techniques of evaluation and analysis, and statistical calculations [can be added as new topics]’. (Moderate.T.2)

Moderate.T.5: “Micro-macro eğitimle birlikte stajyer öğretmenlere, sınavların en azından bir bolumu hazırlatılarak, ölçme değerlendirme süreci mesleğe başlamadan deneyimletilmeli.”

‘Before teacher trainees start teaching profession, they should be provided with experience [practice] in terms of assessment process by making them to prepare at least a part of a language exam during micro-macro training [practicum courses].’ (Moderate.T.5)

Low.T.10: “Ders içeriğinde listening, writing, reading, speaking ve grammar alanlarında ölçme ve değerlendirme yöntemleri gösterilip bol bol uygulama yapılmalıdır. Ayrıca bu derste her bir öğrenci her alan için sınavlar hazırlayıp bunları gerçek bir öğrenci kitlesine uygulamalı ve bu sınavların değerlendirmesini yapmalıdırlar. Ders saati arttırılmalıdır. Ders iki dönem boyunca verilmeli, birinci dönem teknik bilgiler; sınavların nasıl hazırlanıp değerlendirileceği öğretilip, ikinci dönem ise uygulaması yapılmalıdır. 4.sınıfta staj uygulamasında stajyerlerin öğrencilerine öğrettiği konuları her bir alanın (listening, writing, reading, speaking ve grammarin) ölçme ve değerlendirilmesini 2. dönem verilen ölçme ve değerlendirme dersi için yapabilirler. Böylece tüm öğrencilerin verdiğimiz bilgileri ne ölçüde öğrendiklerini anlamış oluruz. Ayrıca öğrenmiş olduğumuz ölçme ve değerlendirme bilgilerini gerçek bir kitlede uygulamış oluruz.”

‘Various different testing and assessment methods for listening, writing, reading, speaking and grammar as well as the practices of them should be given in the [ELTE] course. In addition, students should prepare test items accordingly and should have an opportunity to apply those items and evaluate them. The hours allocated to the course

should be increased. The course should be given for two semesters: In the first semester, the theoretical side of assessment should be taught, and in the second semester, the practical side should be applied. Also, the practice of assessment for each skill can be integrated into the practicum in the second semester. Hence, students will be able to practice assessment in the real classrooms.' (Low.T.10)

Low.T.10: "Grammar, listening ve reading bilgisini ölçecek temel bilgileri edindim. Yalnız speaking ve writing konusunda ders bana yetersiz geldi..... Özellikle speaking ve writing alanlarında ölçme ve değerlendirme yapamıyorum. Speaking ve writing öğretiminde üniversitede aldığım bilgiler çok yetersiz geldi çünkü bunlarla ilgili çok fazla pratik yapma imkânımız olmadı."

'I think I acquired the essential knowledge about assessing grammar, listening and reading. But I think the [ELTE] course was not successful enough in terms of the topics about assessing writing and speaking. [...] Especially, I cannot assess students' speaking and writing skills. In my opinion, the knowledge about assessing speaking and writing skills that I gained at university is inadequate. It is because there was not much practice about that topic.' (Low.T.10)

Moderate.T.12: "Evet, düşünüyorum çünkü görev yapan öğretmenler belirli bir süre sonra kesinlikle ölçme ve değerlendirme kriterlerini göz önünde bulundurmadan sınavlar hazırlıyorlar. Bunun için değişen sisteme uyumlu olarak hizmet içi semineler sürekli yapılmalıdır ve üniversitede öğrenilen bilgiler her an kullanılmak üzere taze tutulmalıdır."

'I think in-service training [about ELTE] is necessary because in-service teachers prepare language exams without paying attention to some assessment criteria after a while. Therefore, constantly, in-service seminars should be provided suitable to the changing needs of the system and what is learned at university should be kept fresh all the time.' (Moderate.T.12)

Low.T.13: "Olabilir. Farklı okul türlerinde ve farklı becerileri ölçmekte karşılaşılan sorunlar üzerine olabilir."

'There might be in-service training. Specifically, it may be about assessing different language skills in different language school contexts as well as the difficulties encountered while assessing.' (Low.T.13)

High.T.19: "Kesinlikle verilmeli. Her okulda bu derse aynı özenin verilmediğini düşünüyorum. Bu kanıya diğer meslektaşlarımdan hazırladığı materyalleri göre vardım."

'In-service training [about ELTE] should be provided. I think each university does not put emphasis on this topic in the same way. I have come to this point on the basis of the materials that are prepared by my colleagues.' (High.T.19)

Moderate.T.6: "Kesinlikle gerekmektedir. Çalıştığım kurumda hiç ölçme ve değerlendirme dersi almamış olan öğretmen arkadaşlar var. Öğretmenlerin, özellikle uygulama sınavı yapılan derslerin öğretmenlerinin eksiklerinin kalmaması gerekir."

'In-service training [about ELTE] is absolutely necessary. There are the colleagues who did not have any training about language assessment at my school. Teachers, especially those who make practical exams in their lessons, should be equipped with such knowledge and skills.' (Moderate.T.6)

High.T.8: "Evet. Örneğin sınavlarımı hazırlarken soruların orta güçlükte olmasına dikkat ediyorum. Doğru- yanlış sorularında cevaba yönelik ipucu vermekten kaçınıyorum. Sorunun açık ve anlaşılır olmasına dikkat ediyorum. Her maddeyi

yalnızca bir beceriyi ölçecek şekilde hazırlıyorum. Boşluk doldurma sorularında boşlukların eşit boyutta olmasına özen gösteriyorum.”

‘It did work [ELTE course was helpful at work]. For example, I pay attention to the difficulty level of test items while preparing exams. I avoid giving clues about correct answer in true-false test items. I try to be careful about the clarity of questions. I prepared each test item which measures only one skill. I pay attention to the length of the gaps in the fill-in-the-blanks test items.’ (High.T.8)

Moderate.T.22: “Evet. Bulunduğum lise de seviye çok düşük hiçbir şekilde hazır ölçme araçları kullanamıyorum. Sorularımın hepsini kendim yazıyorum. Çünkü öğrencilerin ciddi seviye sorunu var.”

‘I have some difficulties in assessing English. The proficiency level is very low in my high school where I work, and thus, I cannot use ready-made tools. I write all my test items on my own because students have problems of critical proficiency levels.’ (Moderate.T.22)

Moderate.T.3: “Aslında bu ders öğretmenlik için çok iyi, sonuçta ülkemiz için yazılı yoklamalar önemli yere sahip.”

‘In fact, this [ELTE] course is very good for teaching profession. All in all, written examinations have an important place in our country [Turkey].’ (Moderate.T.3)

High.T.19: “Türkiye için konuşacak olursak, anadili İngilizce olmayan öğrencilere, günlük hayatlarında duyma ve kullanma şansları olmayan bu dili sınıf ortamında ne kadar verebildiğimiz ve bunun ölçülmesi oldukça elzemdir.”

‘If we talk about Turkey, it is crucial to assess to what extent we can teach English in the classroom to those who are mother-tongue is not English, and also do not have any chance to hear and use English in their daily lives.’ (High.T.19)

Moderate.T.20: “Pratik ve uygulama kısmında zorluk yaşıyorum. Teorik bilğim var fakat devlet okullarındaki ders işleme ve ölçme sistemiyle teorik bilgilerimi birleştirmede zorluk yaşıyorum.”

‘I have difficulties in practice and implementation. I have theoretical knowledge [about language assessment], but I have difficulty in combining such knowledge with the teaching style and testing system at state schools [in Turkey].’ (Moderate.T.20)

The Quotations of Pre-service ELT Teachers (BEFORE)

Before.Pre.23: “Evet. Öğretmen olmak sadece ders anlatmak değildir. Öğrenciye yönelik sınavlar, testler hazırlamak da öğretmenin belki de en çok dikkat etmesi gereken noktadır. Bu yüzden her öğretmen adayının ölçme ve değerlendirme bilgisine sahip olması gerekir.”

‘Assessment in ELT is necessary. Being a teacher is not just about teaching. One of the most important points that every teacher has to pay attention is to prepare language exams suitable for their students. Therefore, every teacher candidate needs to be competent in language assessment.’ (Before.Pre.23)

Before.Pre.77: “Bu dersin önemli ve verilmesi gereken bir ders olduğunu düşünüyorum. Bir öğretmen için nasıl öğreteceği, öğretim hizmetinin kalitesini ve öğretim etkinliklerinin ne kadar başarılı olduğunu belirlenen amaçlara ne kadar hizmet

ettiğini ölçebilmeyi de bilmesi gerektiği ve bu konuda yeterli donanıma sahip olması gerektiğine inanıyorum.”

‘I think this course is necessary and should be given. I believe that every teacher needs to know how to teach, how to evaluate the quality of their teaching, how to assess to what extent their teaching activities are successful, and how to test how much of the goals of their teaching activities are achieved. Teachers should be well-equipped with that kind of knowledge and skills.’ (Before.Pre.77)

Before.Pre.49: “Beklentim kötü testler görmekten çok iyi test örnekleri görmek. Zaten hayatımız boyunca kötü örnekler gördük ve hala görmeye devam ediyoruz. Başarılı örnekler görüp kafamda onların yer edinmesi gerektiğini düşünüyorum.”

‘My expectation from this course is to study on good test samples rather than poor test items. We have already come across poor tests, and still we have seen them. I think it is necessary to study on good test items to construct a good perception of assessment.’ (Before.Pre.49)

Before.Pre.19: “Sınav hazırlama, uygulama, değerlendirme konusunda yeterli bilgi ve beceriyi elde etmeme yardım etmek.”

‘I believe this course will help me to obtain adequate knowledge and skills about preparing, administering, and evaluating tests.’ (Before.Pre.19)

Before.Pre.28: “Bu derste öğrenciyi nasıl sınamamız gerektiğini öğreneceğimizi düşünüyorum. Dersten beklentim nispeten kalıcı olması ve ileriki öğretmenlik hayatımda kullanacağım bilgileri edineceğim bir süreci geçirmek.”

‘In this course, I think we will learn how to assess language learners. I expect the course will be relatively permanent, and I want to gain the knowledge that I will use in my professional teaching life.’ (Before.Pre.28)

Before.Pre.57: “Metodoloji derslerinden de öğrendiklerimiz harmanlayarak öğretmenlik mesleğine gerek olan bilgisi gerekse yöntem bilgisi konusunda deneyimli bir şekilde ders vereceğimi düşünüyorum.”

‘[At the end of this course], I think I will be able to teach and assess properly by combining teaching methodology knowledge as well as professional knowledge that I learned [throughout training years]. (Before.Pre.57)

Before.Pre.150: “Almış olduğumuz 4 yıllık eğitim boyunca nerelerde bize gerçek anlamda testler uygulandı ya da uygulanmadı? Bunları görüp süzgeçten geçirebileceğim bilgileri öğrenmiş olacağımı düşünüyorum.”

‘Throughout 4-year teacher training, have all the exams we have taken so far been implemented correctly or not? I think I will learn some kind of knowledge to help to evaluate such an issue.’ (Before.Pre.150)

Before.Pre.50: “Sınav hazırlama konusunda bilgili bir öğretmen olacağımı düşünüyorum. Aynı zamanda öğrettiklerim ile sorduklarım arasında bağlantı kurabilecek ve sınavların öğrenci psikolojisi üzerine olan etkilerini görebileceğim.”

‘I think I will be a knowledgeable teacher about language assessment. In addition, I will be able to make a connection between what I teach and assess as well as to see the psychological effects of tests on students [backwash effect].’ (Before.Pre.50)

Before.Pre.111: “Son sınıf olduğum için öncelikli beklentim dersten kalmamak ve bir şeyler öğrenebilmek.”

‘Since I am a senior student, my first expectation is to pass the course and to learn something [about assessment]. (Before.Pre.111)

Before.Pre.7: “Kesinlikle verilmesinin gerekli olduğunu düşünüyorum. Eğitim fakültesinde okuyan bir öğretmen adayı için kendi alanıyla ilgili nasıl değerlendirme yapacağını, değerlendirme yaparken gerekli olan ölçütlerin neler olduğunu derinlemesine öğrenmesi kaliteli ve başarılı bir mesleki yaşantı açısından gereklidir.”

‘I think this course should be given. It is necessary for pre-service teachers studying at faculty of education in order to learn and know how to assess language skills and the evaluation criteria for a successful and qualified professional life.’ (Before.Pre.7)

Before.Pre.35: “Dersin verilmesi gerektiğini düşünüyorum çünkü şimdiye kadar nasıl öğreneceğimiz gördük. Bu son sene de öğrendiklerimizi nasıl ölçeceğimizi öğreneceğiz. Ne öğrendik, nereye geldik, öğretmen olunca değerlendirme yapacak seviye geldik mi? Gibi sorulara cevap olabilir.”

‘I think this course should be taught because so far, we have learned how to teach. At this last year, we will learn how to test. What did we learn, which level are we at, and have we come to the level of making evaluations? This course may answer such questions.’ (Before.Pre.35)

Before.Pre.165: “Faydalı olacağına inanıyorum. Testinge bakış açımızı değiştirecek ve önyargılarımızı kırarak nasıl iyi bir eğitim verebiliriz onu anlamış olacağız.”

‘I believe this course will be beneficial. It will change our point of view towards testing and we will understand how to give a better education by breaking down our prejudices.’ (Before.Pre.165)

Before.Pre.58: “Sınav hazırlama ve değerlendirme eğitim-öğretim sürecinin en önemli basamaklarından birisi bence. Hem öğrenci hem de öğretmen için dönüt niteliği taşıyor. Öğrenci yanlışlarını görüp doğrusunu öğrenirken, öğretmen de tekniğini değerlendirmiş oluyor. Ayrıca Türkiye’de eğitim sisteminin her basamağı sınav olduğu için öğretmen nasıl sınav hazırlaması gerektiğini bilmeli. O yüzden bu dersi önemli ve gerekli buluyorum.”

‘In my opinion, testing and evaluation are one of the most important stages of educational process. Assessment is a feedback both for teachers and students. While students have the chance to see their mistakes and learn the correct ones, teachers evaluate their teaching techniques [as a result of assessment]. Furthermore, since every stage of educational system in Turkey is dependent on exams, every teacher should know how to prepare tests. Therefore, this course is important and necessary.’ (Before.Pre.58)

Before.Pre.104: “Bu ders önemlidir çünkü Türkçe olarak aldığımız ölçme ve değerlendirme dersi genel anlamda her öğretmenlik alanı için gerekli. Fakat biz tamamen yeni bir dil öğretiyoruz ve genel uygulamaların dışında İngilizce’yi ölçmenin daha farklı yolları olmalıdır.”

‘This course is important because the course of assessment and evaluation instructed in Turkish [at the third year of training as general educational and pedagogical course] is very general. But we teach a completely different language [English as a foreign language] and apart from general practices, there must be different ways to assess English language.’ (Before.Pre.104)

Before.Pre.63: “Tabii ki bir öğretmen için oldukça gerekli bir ders. Ancak olaya Türkiye şartlarında bakarsak her öğrenci bu dersi alıyor mu, alıyorsa ne kadar etkili alıyor ve daha da önemlisi öğretmen olunca hangi öğretmen bu dersin gerekliliklerini yerine getiriyor? Oldukça içler acısı bir durum olduğunu düşünüyorum.”

'Sure, it is a necessary course for every teacher. But when we consider the conditions of Turkey, some questions arise: Has each pre-service teacher taken this course?; if they have taken, how effective has the course been?; after graduation, have they fulfilled the requirements of this course in their teaching? I think it is an awful condition.' (Before.Pre.63)

The Quotations of Pre-service ELT Teachers (AFTER)

After.High.Pre.7: “Bir öğretmenın sahip olması gereken ve bence en önemli şeylerden biri sınav hazırlamaktır. Önemli olan sadece öğretmek değil aynı zamanda ölçmektir. Bu ders sayesinde bunu doğru bir şekilde yapabileceğime inanıyorum.”

'One of the most important competencies that every teacher must possess is preparing language exams. Teaching is not the only important thing, but at the same time, testing/assessment is very important. I believe we will be able to assess properly thanks to this course.' (After.High.Pre.7)

After.Moderate.Pre.113: “Ölçme-değerlendirme dersi aslında şu ana kadarki öğrenmelerimizin tamamlayıcısı niteliğinde. Çünkü öğrettiğimiz şeyleri test etmek hem öğrencilerin gelişimini sağlamak hem de öğrenme sürecini tamamlamak açısından çok önemli.”

'In fact, this course is complementary to what we have learned so far [throughout our teacher training]. It is because it is very important to test what we teach, to assess the progress of students' language development, and to supply learning/teaching process.' (After.Moderate.Pre.113)

After.Moderate.Pre.100: “Orta olduğumu düşünüyorum. Çünkü burada aldığım teorik bilgileri uygulamada ne kadar etki kullanacağımı bilemiyorum, ilerleyen zamanlarda tecrübeyle birlikte yükseltebileceğimi düşünüyorum.”

'I perceived myself being qualified in language assessment at a moderate level. It is because I cannot foresee how effectively I will put theoretical knowledge into practice. I believe I will be better in language assessment via experience.' (After.Moderate.Pre.100)

After.High.Pre.90: “Kendimi daha yeterli görüyorum. Sınav veya aktivite hazırlarken yaptığım yanlışları gördüm ve nasıl düzeltebileceğimi öğrendim.”

'I perceive myself more competent now [at the end of the semester]. I noticed my mistakes while preparing language test items or activities, and I learned how to fix them.' (After. High.Pre.90)

After.Low.Pre.78: “Deneyimimiz çok olmadığı için bu konuda kendimi yeterli görmüyorum. Ama tabii ki mesleğe başladıktan sonraki süreçte bunun değişeceğini umuyorum. Deneyerek görerek daha yeterli hale geleceğimize inanıyorum.”

'I do not perceive myself being competent in language assessment because of the lack of experience. But I hope this situation will change after starting my career as an English teacher. I believe I will be better by means of practicing.' (After.Low.Pre.78)

After.Moderate.Pre.100: “Ders sonunda ölçme değerlendirmeyi dilin her alanında (grammar, vocab, listening, ...) nasıl yapacağımı, bunu yaparken nelere dikkat etmem gerektiğini öğrendim.”

'At the end of the course, I learned how to assess language skills (grammar, vocabulary, listening, etc.) and also, the criteria which I should bear in mind while assessing.' (After.Moderate.Pre.100)

After.Moderate.Pre.119: “Sınav hazırlarken dikkat etmemiz gereken şeyleri öğrendik. Amacımıza sınıfımızın düzeyine göre sınavları nasıl oluşturmamız gerektiğini. Bu sınavları daha etkili ve güvenilir yapabilmek için kullanacağımız yöntemleri öğrendik.”

‘We received information about the criteria that we have to pay attention while designing language tests. Also, we learned how we can prepare language exams according to the level of our classes as well as our teaching goals. Besides, we studied on the procedures which can be utilized to increase the reliability and validity of language tests.’
(After.Moderate.Pre.119)

After.Moderate.Pre.44: “Öğretme ve değerlendirme arasında sıkı bir bağın olduğunu. Test tekniklerini, değerlendirmeyi, bir ölçme aracında olması gereken özellikleri, dilbilgisi, kelime, okuma, yazma, dinleme gibi becerilerle ilgili nasıl soru hazırlanacağını öğrendim.”

‘I learned at the end of this course that there is a close relationship between teaching and assessment. I also learned about language testing techniques, evaluation, the features that any measurement tool should possess, and how to prepare test items for language skills such as grammar, vocabulary, reading, writing, and listening.’
(After.Moderate.Pre.44)

After.Moderate.Pre.127: “Bir öğretmen adayı olarak ileride uygulayacağım testler için neleri göz önünde bulunduracağımı, nasıl test oluşturup değerlendirebileceğimi, dönüt vereceğimi, öğretmesi açısından faydalı oldu.”

‘As a language teacher candidate, I learned about what sort of aspects that I have to consider while designing language tests, how to prepare and evaluate test items, and how to give feedback in my future teaching career. That is why, this course was beneficial.’ (After.Moderate.Pre.127)

After. Moderate.Pre.31: “Evet karşıladı. Üniversitede 4 sene boyunca İngilizce öğretimi ile ilgili dersler aldık yani sınıf içi etkinliklerinin nasıl düzenlememiz gerektiğini öğrendik. Ancak ders sonunda bunları nasıl değerlendireceğimizi bilmiyorduk. Bu açıdan bu ihtiyacı karşıladığını düşünüyorum.”

‘This course met my expectations. Throughout our 4-year-university life, we took the courses about how to teach English language; I mean, how to organize language teaching activities in the class. But we had not known how to assess those at the end of those courses. From this point, this course filled this gap.’ (After. Moderate.Pre.31)

After.Moderate.Pre.87: “Evet. Dersin Türkçesini 3. sınıfta almıştım fakat kendi alanıma odaklı bu dersi almak İngilizce öğretmenliği kariyerimde daha etkili bir testing ve evaluation aşamasını gerçekleştirmiş olacağım.”

‘This course was helpful. I took a similar version of this course in the third year, but it was Turkish and general [Measurement and Evaluation in Education]. Thus, by taking this specific course dedicated to English language testing and evaluation, I believe I will be more effective in terms of language assessment in my professional life.’
(After.Moderate.Pre.87)

After.Moderate.Pre.83: “Bence yeterliydi. Becerilerin detaylarını vermeden önce teknik bilgileri vermesi bizim konular hakkında zemin hazırlamamızı sağladı ve bu güzeldi.”

‘I think the course content was sufficient. Providing detailed theoretical and practical knowledge about assessing language skills made us understand the concept of assessment, which was pretty good.’ (After.Moderate.Pre.83)

After.Moderate.Pre.112: “Bence dersin içeriği yeterli çünkü her alan ve beceri için yeteri kadar alternatif var öğrencinin neye ihtiyacı olduğunu öğrenebiliyoruz.”

‘I think the course content was sufficient because there were adequate options for assessment examples for each language area and skill, and also, we learned what students would need as a result of assessment.’ (After.Moderate.Pre.112)

After.Moderate.Pre.100: “Bu ders beklentilerimi ve ihtiyaçlarımı karşıladı. Öğretmenin ders anlatımında iyi olması, açık, anlaşılır, net bir şekilde anlatması sayesinde yüksek verim aldığımı düşünüyorum.”

‘The course responded to my needs and expectations. I think I had learned a lot because of our course lecturer whose teaching was very good, clear, comprehensible and explanative.’ (After.Moderate.Pre.100)

After.Low.Pre.4: “Yeterince karşılamadı çünkü dönemin tamamı teorik ders ile geçti. Uygulama kısmı ödev olarak verildi. Teorik kısımları önceden okuyup gelebilecekken sınıfta buna ağırlık verildi ve uygulama kısmında hiç deneyimimiz olmadığı için sorun yaşayabileceğimi düşünüyorum.”

‘The course did not respond to my needs adequately because the whole semester was theoretical. The practical part was given only as an assignment. While the theoretical parts could have been read at home, there was much more emphasis on that side in the lessons. Therefore, since we had no experience in terms of practicing assessment, there will be problems in the future.’ (After.Low.Pre.4)

After.Moderate.Pre.30: “Evet, içerik olarak yeterliydi. Ama kullandığımız coursebookta typelar biraz eskiydi. Günümüze göre geliştirilebilir. Onun dışında faydalı ve yeterli bir ders.”

‘Yes, the course content was sufficient. But the types [examples] in the coursebook we used in the lessons were a bit out of date. They may be updated. Apart from this, the course is beneficial and adequate.’ (After.Moderate.Pre.30)

After.Moderate.Pre.49: “Kısmen karşıladı. Öğretmenin etkin olduğu yerlerde daha verimli geçti fakat en çok öğrencilerin sunum yaptığı yerlerde dersi anlamakta daha zorlandım.”

‘The course partly responded to my needs. It was more efficient when the course lecturer was active in the class but for me, it was too hard to understand when the course takers did presentations about the topics.’ (After.Moderate.Pre.49)

After.Moderate.Pre.95: “İçerik açısından yeterli fakat son sınıfa bırakılmaması gereken bir ders olduğunu düşünüyorum. Daha önce alınıp son sınıfta uygulama şansına olabilirdi.”

‘I think the course content was sufficient, but the problem is that the course is given too late [in the last semester]. If I had taken the course before, there would have been more chance to practice it in the last semester.’ (After.Moderate.Pre.95)

After.Moderate.Pre.110: “Dersin içeriği yeterliydi. Fakat öğrencilere konular verilerek sınav hazırlanmaları ya da mini quiz tarzından bu dersi pratiğe dökebilecekleri bir ödev hazırlattırabilirdi.”

‘The course content was sufficient. However, the course takers would have been made to prepare exams according to the topics or to design mini quizzes in order to practice assessment.’ (After.Moderate.Pre.110)

After.Moderate.Pre.26: “Teorik açıısından yeterli olduğunu düşünüyorum ama pratikte ve pratik kullanımı hakkında yeterli olmadığını düşünüyorum. En azından stajın da bu dersi kapsamısı gerektiğı düşüncesindeyim.”

‘The course was sufficient in terms of its theoretical knowledge, but I think it was inadequate in terms of its practical side. I think the course should be integrated into the practicum.’ (After.Moderate.Pre.26)