

**THE DESCRIPTION OF PRE-SERVICE EFL TEACHERS'
LANGUAGE ASSESSMENT LITERACY DEVELOPMENT PROCESS**

Murat Şükür

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

**GAZİ UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES**

SEPTEMBER, 2023

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 12 ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Murat
Soyadı : Şükür
Bölümü : İngiliz Dili Eğitimi
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı : Hizmet Öncesi İngilizce Öğretmenlerinin Dilde Ölçme-Değerlendirme Okuryazarlığı Geliştirme Sürecinin Betimlenmesi

İngilizce Adı : The Description of Pre-service EFL Teachers' Language Assessment Literacy Development Process

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Murat ŞÜKÜR

İmzası:

JÜRİ ONAY SAYFASI

Murat Şükür tarafından hazırlanan “The Description of Pre-service EFL Teachers’ Language Assessment Literacy Development Process” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi İngiliz Dili Eğitimi Anabilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Gonca YANGIN EKŞİ

İngiliz Dili Eğitimi Anabilim Dalı, Gazi Üniversitesi

Başkan: Doç. Dr. Asuman AŞIK

İngiliz Dili Eğitimi Anabilim Dalı, Gazi Üniversitesi

Üye: Doç. Dr. Burçak YILMAZ YAKIŞIK

İngiliz Dili Eğitimi Anabilim Dalı, Gazi Üniversitesi

Üye: Dr. Öğretim Üyesi İsmail Fırat ALTAY

İngiliz Dili Eğitimi Anabilim Dalı, Hacettepe Üniversitesi

Üye: Prof. Dr. Nuray ALAGÖZLÜ

İngiliz Dili Eğitimi Anabilim Dalı, Hacettepe Üniversitesi

Tez Savunma Tarihi: 22/08/2023

Bu tezin İngiliz Dili Eğitimi Anabilim Dalında Doktora tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban ÇETİN

Eğitim Bilimleri Enstitüsü Müdürü

To my wife, Sinem and my daughter, Rüya



ACKNOWLEDGEMENTS

First of all, I would like to express my sincere thanks to my supervisor, Prof. Dr. Gonca YANGIN EKŞİ. I was always able to get answers to my questions easily and quickly from her. Therefore, it would not have been possible to finish this dissertation without her endless efforts.

I would also like to thank Assoc. Prof. Dr. Asuman AŞIK and Assist. Prof. Dr. İsmail Fırat ALTAY for their useful feedback in the thesis monitoring committees. In addition, my special thanks go to Prof. Dr. Nuray ALAGÖZLÜ and Assoc. Prof. Dr. Burçak YILMAZ YAKIŞIK for their comprehensive and attentive feedback and for their kindness in taking time from their holidays to attend the thesis defense.

I would like to express my sincere thanks to Dr. Tuba KARAGÖZ DİLEK and Dr. Erdem KARATAŞ, who were always encouraging and helped me find answers by listening to the problems sincerely. Many thanks to Assist. Prof. Dr. Yusuf Emre YEŞİLYURT and Dr. Samed Yasin ÖZTÜRK, Dr. Serhan KÖSE, Dr. Sevinç ERGENEKON and Dr. Gamze EMİR, with whom I enjoyed sharing the same room at Gazi University, for their sincere collegueship. I would like to thank our esteemed department chair, Hamide ÇAKIR SARI, who has always treated us sincerely and whose support we always felt at Niğde Ömer HALİSDEMİR University. I sincerely hope that he will recover as soon as possible and join us again with his professional love and sincerity.

I would like to extend my endless and most sincere thanks and love to my life mate Sinem, who has always been with me in my most difficult days and who has faced every difficulty with me, for her endless life energy, calmness, love, and immaculate heart. Endless love to my daughter Rüya, my most precious one, who has added color to our lives since the first day we learned that she would be born and made us experience the feeling of motherhood

and parenthood. I hope we will realize the things I have been dreaming of for years together on very beautiful days, my dear daughter. I would like to convey my respect and love to my mother-in-law, Döndü KEÇECİ, for her endless help, motherhood, and sincere love in the difficult process before and after our daughter was born. I am also grateful to my dear sister-in-law Didem HANCI and brother-in-law Uğur HANCI for their support of me and my wife during this difficult process.

I would also like to extend my sincerest love to my dear mother Nazan, my dear father Sefa, and my dear sisters, who have always contributed to me getting to where I am today. Thank you, my dear extended family. I am always grateful to my dear grandfather Ali, whose illness we learned about during my thesis writing process. Now I live with his memories. I will always miss you. You are always in my heart, Paşam. Last but not least, I would like to express my love to my dear aunt Esme YILMAZ, who passed away very shortly before the defense of this thesis. Rest in peace, my dear one.

HİZMET ÖNCESİ İNGİLİZCE ÖĞRETMENLERİNİN DİLDE ÖLÇME-DEĞERLENDİRME OKURYAZARLIĞI GELİŞTİRME SÜRECİNİN BETİMLENMESİ

(Doktora Tezi)

Murat Şükür
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eylül, 2023

ÖZ

Bu araştırma, İngilizce öğretmeni adaylarının dilde ölçme-değerlendirme okuryazarlığı (DÖDO) konusundaki gelişim süreçlerini betimlemeyi amaçlamaktadır. Bu amaçla, çoklu vaka çalışması yürütülmüştür. Katılımcılar, Gazi Üniversitesi İngilizce Öğretmenliği bölümünde dört farklı sınıf düzeyinde öğrenim gören yirmi öğretmen adayından oluşmaktadır. Bunlardan dört tane birinci sınıf ve beş tane ikinci sınıf öğrencisi birinci vakayı oluşturmuştur. Öte yandan, ikinci vaka altı tane üçüncü sınıf İngilizce öğretmeni adayından oluşurken, üçüncü vaka ise beş tane son sınıf İngilizce öğretmeni adayını içermektedir. Bu katılımcılardan, DÖDO'nun üç ana bileşeni olan bilgi, ilkeler ve becerilere ilişkin nitel veriler toplanmıştır. Veri toplama araçları, görüş günlükleri, üç tur yarı yapılandırılmış birebir ve odak grup görüşmeleri, yabancı dil değerlendirme senaryoları, ders gözlemleri ve izlenceler, ders planları ve ödevler gibi dokümanlardan oluşmuştur. Tüm veriler araştırmacı tarafından nitel içerik analizine tabi tutulmuştur. Ayrıca, kodlayıcılar arası güvenilirliği sağlamak için başka bir kodlayıcı da verilerin rastgele seçilen belirli bir kısmını analiz etmiştir. DÖDO'nun bilgi bileşenine ilişkin bulgular, daha önce değerlendirme konusunda herhangi bir eğitim almamış olan Vaka 1'deki katılımcıların temel terimler ve değerlendirme türlerine ilişkin bazı önemli kavram yanılgılarına sahip olduklarını ortaya koymaktadır. Ancak bu bulgu, Eğitimde Ölçme Değerlendirme ve İngilizce Öğretiminde Ölçme Değerlendirme derslerinden birini veya her ikisini almış

olmalarına rağmen Vaka 2 ve Vaka 3'teki katılımcılar için de çoğunlukla geçerlidir. DÖDO'nun beceri bileşeniyle ilgili bulgular ise, öğretmen adaylarının çocuk yaşta İngilizce öğrenen öğrencilerin alıcı becerilerini değerlendirmede zorluklar yaşadığını göstermektedir. Bunun aksine, öğretmen adayları hem çocuk hem de ergen yaşta İngilizce öğrenenlerin üretken becerilerini değerlendirmede yüksek düzeyde okuryazarlığa sahiptirler. İlkeler bileşenine ilişkin olarak İngilizce öğretmeni adaylarının etik bir değerlendirme sürecinin özelliklerinin farkında oldukları bulunmuştur. Ayrıca, dilde ölçme-değerlendirme sürecinde etiği sağlama konusundaki okuryazarlıkları bazı zayıf ve güçlü yönler içermektedir. Bu bulgular ışığında, Gazi Üniversitesi İngilizce Öğretmenliği bölümünde öğrenim gören öğretmen adaylarının, öğretmen yetiştirme programının hedeflediği şekilde DÖDO'larını bazı hususlar bakımından yeterince geliştiremedikleri söylenebilir. Özellikle, dilde ölçme-değerlendirmenin teorik yönleri ve çocuk yaşta İngilizce öğrenen öğrencilerin dil becerilerine ve bileşenlerine ilişkin yeterlilik ve performanslarının nasıl değerlendirileceği konusunda okuryazarlıklarını geliştirmeleri gerekmektedir.



Anahtar Kelimeler : Dilde ölçme-değerlendirme okuryazarlığı, bilgi, beceriler, prensipler, öğretmen eğitimi

Sayfa Adedi : xx + 50

Danışman :Prof. Dr. Gonca YANGIN EKŞİ

**THE DESCRIPTION OF PRE-SERVICE EFL TEACHERS’
LANGUAGE ASSESSMENT LITERACY DEVELOPMENT PROCESS
(Doctoral Dissertation)**

Murat Şükür

**GAZİ UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES**

September, 2023

ABSTRACT

This research aims to describe the pre-service English teachers’ development process in language assessment literacy (LAL). To this end, the present multiple-case study was conducted. The participants were twenty teacher trainees studying at four grade levels of the Gazi University ELT department. Of these, four freshmen and five sophomores constituted the first case. On the other hand, the second case consisted of six third-year teacher candidates, while the third one involved five senior pre-service English teachers. From these participants, qualitative data were collected regarding the three main components of LAL: knowledge, principles, and skills. The data collection tools were opinion journals, three rounds of semi-structured one-to-one and focus group interviews, scenarios for foreign language assessment, lesson observations, and documents like syllabuses, lesson plans, and assignments. All the data were content analyzed by the researcher iteratively. Moreover, another coder also analyzed a randomly selected sample of the data in order to ensure inter-coder reliability. The findings related to the knowledge component of LAL reveal that the participants in Case 1, who do not have any previous training in assessment, have some significant misconceptions regarding the key terms and types of assessment. However, the participants in Cases 2 and 3 also do not have sufficient knowledge of basic concepts despite the fact that they took one or both the Measurement and Evaluation in Education and Measurement and Evaluation in English Language Teaching courses. Regarding the skills component of LAL, the findings indicate that the teacher candidates have challenges in

assessing the receptive skills of young language learners. In contrast, they have a high literacy level in assessing the productive skills of young and adolescent English learners. As for the principles component, the present study found that pre-service English teachers are aware of the features of an ethical assessment process. Furthermore, their literacy in ensuring ethics in the language assessment includes weaknesses and strengths. Considering these findings, the teacher trainees studying at the ELT department of Gazi University need to improve their LAL more as the teacher training program aims. In particular, they need to develop their literacy regarding the theoretical aspects of language assessment and how to assess the young language learners' competence and performance in language skills and components.



Key words: Language assessment literacy; knowledge; skills; principles; teacher education

Page Number: xx + 350

Supervisor: Prof. Dr. Gonca YANGIN EKŞİ

TABLE OF CONTENTS

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI	ii
JÜRİ ONAY SAYFASI.....	iii
ACKNOWLEDGEMENTS	v
ÖZ	vii
ABSTRACT.....	ix
LIST OF TABLES	xvii
LIST OF FIGURES	xx
LIST OF ABBREVIATIONS	xxi
CHAPTER I	1
INTRODUCTION.....	1
Introduction	1
Background to the Study	1
Problem Statement	2
Aim of the Study	4
Significance of the Study	5
Theoretical Framework.....	6
Constructivism	6
Sociocultural Theory.....	7
Definition of the Key Terms	10
CHAPTER II.....	12
LITERATURE REVIEW	12
Assessment	12

Reliability	13
Validity	15
Authenticity	16
Practicality	16
Washback Effect	18
Formative Assessment.....	18
Self-Assessment	21
Peer Assessment	23
Summative Assessment	23
Norm-referenced and Criterion-referenced Assessment	24
Diagnostic Assessment	26
Performance-based Assessment.....	26
Dynamic Assessment	27
Assessment Literacy	27
Language Assessment Literacy	29
Related Studies Conducted in Other Countries	31
Related Studies Conducted in Turkey	34
CHAPTER III	37
METHODOLOGY	37
Research Design.....	37
Participants.....	38
Context.....	41
Data Collection Procedure	42
Semi-structured Interviews	44
Focus Group Interviews.....	46
Opinion Journals.....	47
Lesson Observations	48
Scenarios for Foreign Language Assessment	49
Documents	49
Data Analysis	49
Pilot Study.....	52
Trustworthiness	52
CHAPTER IV.....	57

FINDINGS.....	57
Case 1: First and Second Graders.....	57
Literacy in the Knowledge Component of LAL.....	63
Knowledge about the Key Concepts of Assessment.....	63
<i>Weak Literacy of the Key Concepts of Assessment (First Semi-structured Interview)</i>	69
<i>Strong Literacy of the Key Concepts of Assessment (First Semi-structured Interview)</i>	72
<i>Weak Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)</i>	73
<i>Strong Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)</i>	78
Knowledge about the Types of Assessment	80
<i>Weak Literacy of the Types of Assessment (First Semi-structured Interview)</i>	80
<i>Strong Literacy of the Types of Assessment (First Semi-structured Interview)</i>	86
<i>Weak Literacy of the Types of Assessment (Third Semi-structured Interview)</i>	87
<i>Strong Literacy of the Types of Assessment (Third Semi-structured Interview)</i>	91
Awareness of the Factors Affecting the Quality of Language Assessment.....	93
Literacy in the Skills Component of LAL	96
Assessing the Language Skills.....	96
<i>Assessing Listening.....</i>	97
<i>Assessing the Listening Skills of Young Language Learners</i>	97
<i>Assessing the Listening Skills of Adolescent Language Learners</i>	102
<i>Assessing Speaking Skills.....</i>	105
<i>Assessing the Speaking Skills of Young Language Learners.....</i>	105
<i>Assessing the Speaking Skills of Adolescent Language Learners.....</i>	110
<i>Assessing the Reading Skills</i>	114
<i>Assessing the Reading Skills of Young Language Learners.....</i>	114
<i>Assessing the Reading Skills of Adolescent Language Learners.....</i>	117
<i>Assessing the Writing Skills</i>	120
<i>Assessing the Writing Skills of Young Language Learners.....</i>	120
<i>Assessing the Writing Skills of Adolescent Language Learners.....</i>	123
Assessing the Language Components: Young Language Learners	126
Assessing the Language Components: Adolescent Language Learners.....	136
Scoring, Interpreting, and Using the Assessment Results	144
Literacy Related to the Principles Component of LAL	156
Suggestions and Expectations for the Improvement of LAL.....	176
Case 2: Third Graders.....	177

Literacy in the Knowledge Component of LAL.....	185
<i>Knowledge about the Key Concepts of Assessment</i>	185
<i>Weak Literacy of the Key Concepts of Assessment (First Semi-structured Interview)</i>	190
<i>Strong Literacy of the Key Concepts of Assessment (First Semi-structured Interview)</i>	193
<i>Weak Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)</i>	195
<i>Strong Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)</i>	196
<i>Knowledge about the Types of Assessment</i>	198
<i>Weak Literacy of the Types of Assessment (First Semi-structured Interview)</i>	203
<i>Strong Literacy of the Types of Assessment (First Semi-structured Interview).....</i>	204
<i>Weak Literacy of the Types of Assessment (Third Semi-structured Interview)</i>	206
<i>Strong Literacy of the Types of Assessment (Third Semi-structured Interview).....</i>	207
<i>Awareness of the Factors Affecting the Quality of Language Assessment.....</i>	209
Literacy in the Skills Component of LAL	211
<i>Assessing the Language Skills</i>	211
<i>Assessing the Listening Skills.....</i>	211
<i>Assessing the Listening Skills of Young Language Learners</i>	211
<i>Assessing the Listening Skills of Adolescent Language Learners</i>	213
<i>Assessing the Speaking Skills</i>	215
<i>Assessing the Speaking Skills of Young Language Learners.....</i>	215
<i>Assessing the Speaking Skills of Adolescent Language Learners.....</i>	217
<i>Assessing the Reading Skills</i>	219
<i>Assessing the Reading Skills of Young Language Learners</i>	219
<i>Assessing the Reading Skills of Adolescent Language Learners.....</i>	221
<i>Assessing the Writing Skills</i>	224
<i>Assessing the Writing Skills of Young Language Learners.....</i>	224
<i>Assessing the Writing Skills of Adolescent Language Learners.....</i>	225
<i>Assessing the Language Components: Young Language Learners</i>	226
<i>Assessing the Language Components: Adolescent Language Learners.....</i>	229
<i>Scoring, Interpreting, and Using the Assessment Results</i>	232
Literacy in the Principles Component of LAL.....	234
Reflections, Suggestions, and Expectations for the Improvement of LAL.....	243
Case 3: Fourth Graders.....	245
Literacy in the Knowledge Component of LAL	254

Knowledge about the Key Concepts of Assessment	254
Weak Literacy of the Key Concepts of Assessment (First Semi-structured Interview)	259
Strong Literacy of the Key Concepts of Assessment (First Semi-structured Interview)	262
Weak Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)	264
Strong Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)	265
Knowledge about the Types of Assessment	266
Weak Literacy of the Types of Assessment (First Semi-structured Interview)	270
Strong Literacy of the Types of Assessment (First Semi-structured Interview)	272
Weak Literacy of the Types of Assessment (Third Semi-structured Interview)	274
Strong Literacy of the Types of Assessment (Third Semi-structured Interview)	275
Awareness of the Factors Affecting the Quality of Language Assessment.....	276
Literacy in the Skills Component of LAL	277
Assessing the Language Skills	277
Assessing Listening.....	278
Assessing the Listening Skills of Young Language Learners	278
Weak Literacy in Assessing the Listening Skills of Young Language Learners ...	278
Strong Literacy in Assessing the Listening Skills of Young Language Learners..	279
Assessing the Listening Skills of Adolescent Language Learners	279
Weak Literacy in Assessing the Listening Skills of Adolescent Language Learners	280
Strong Literacy in Assessing the Listening Skills of Adolescent Language Learners	280
Assessing Speaking.....	281
Assessing the Speaking Skills of Young Language Learners.....	281
Assessing Reading	282
Assessing the Reading Skills of Young Language Learners.....	282
Assessing the Reading Skills of Adolescent Language Learners.....	284
Assessing Writing	285
Assessing the Writing Skills of Young Language Learners.....	285
Assessing the Writing Skills of Adolescent Language Learners.....	286
Assessing the Language Components: Young Language Learners	288
Assessing the Language Components: Adolescent Language Learners.....	291
Scoring, Interpreting, and Using the Assessment Results	293

Literacy in the Principles Component of LAL.....	295
Reflections on the Improvement of LAL.....	305
CHAPTER V	307
DISCUSSION AND CONCLUSION	307
Discussion of the Findings and Implications.....	307
Recommendations for Further Research.....	313
REFERENCES.....	315
APPENDICES	329



LIST OF TABLES

Table 1. <i>Demographic Information of the Participants</i>	40
Table 2. <i>Comparison of the Old and New Curricula</i>	42
Table 3. <i>LAL Components and the Data Collection Tools</i>	43
Table 4. <i>Data Collection Procedure</i>	44
Table 5. <i>Duration of the Semi-structured Interviews</i>	46
Table 6. <i>Duration of the Focus Group Interviews</i>	47
Table 7. <i>Knowledge about the Key Concepts of Assessment</i>	64
Table 8. <i>Knowledge about the Types of Assessment</i>	81
Table 9. <i>Assessing the Listening Skills of Young Language Learners</i>	98
Table 10. <i>Assessing the Listening Skills of Adolescent Language Learners</i>	102
Table 11. <i>Assessing the Speaking Skills of Young Language Learners</i>	105
Table 12. <i>Assessing the Speaking Skills of Adolescent Language Learners</i>	110
Table 13. <i>Assessing the Reading Skills of Young Language Learners</i>	115
Table 14. <i>Assessing the Reading Skills of Adolescent Language Learners</i>	117
Table 15. <i>Assessing the Writing Skills of Young Language Learners</i>	121
Table 16. <i>Assessing the Writing Skills of Adolescent Language Learners</i>	124
Table 17. <i>Assessing Grammar: Young Language Learners</i>	127
Table 18. <i>Assessing Vocabulary: Young Language Learners</i>	132

Table 19. <i>Assessing Grammar: Adolescent Language Learners</i>	137
Table 20. <i>Assessing Vocabulary: Adolescent Language Learners</i>	140
Table 21. <i>Scoring the Answers or Performances</i>	145
Table 22. <i>Interpreting and Using the Assessment Results</i>	152
Table 23. <i>Awareness of Ethical Assessment Process</i>	157
Table 24. <i>Ensuring Ethical Assessment</i>	162
Table 25. <i>Suggestions and Expectations for the Improvement of LAL</i>	177
Table 26. <i>Knowledge about the Key Concepts of Assessment</i>	186
Table 27. <i>Knowledge about the Types of Assessment</i>	199
Table 28. <i>Awareness of the Factors Affecting the Quality of Language Assessment</i>	209
Table 29. <i>Assessing the Listening Skills of Young Language Learners</i>	211
Table 30. <i>Assessing the Listening Skills of Adolescent Language Learners</i>	214
Table 31. <i>Assessing the Speaking Skills of Young Language Learners</i>	216
Table 32. <i>Assessing the Speaking Skills of Adolescent Language Learners</i>	218
Table 33. <i>Assessing the Reading Skills of Young Language Learners</i>	220
Table 34. <i>Assessing the Reading Skills of Adolescent Language Learners</i>	222
Table 35. <i>Assessing the Writing Skills of Young Language Learners</i>	224
Table 36. <i>Assessing the Writing Skills of Adolescent Language Learners</i>	225
Table 37. <i>Assessing Grammar: Young Language Learners</i>	227
Table 38. <i>Assessing Vocabulary: Young Language Learners</i>	228
Table 39. <i>Assessing Grammar: Adolescent Language Learners</i>	230
Table 40. <i>Assessing Vocabulary: Adolescent Language Learners</i>	231
Table 41. <i>Scoring the Answers or Performances</i>	232
Table 42. <i>Interpreting and Using the Assessment Results</i>	234
Table 43. <i>Awareness of Ethical Assessment Process</i>	235

Table 44. <i>Ensuring Ethical Assessment</i>	238
Table 45. <i>Reflections, Suggestions, and Expectations for the Improvement of LAL</i>	244
Table 46. <i>Knowledge about the Key Concepts of Assessment</i>	255
Table 47. <i>Knowledge about the Types of Assessment</i>	267
Table 48. <i>Awareness of the Factors Affecting the Quality of Language Assessment</i>	276
Table 49. <i>Assessing the Listening Skills of Young Language Learners</i>	278
Table 50. <i>Assessing the Listening Skills of Adolescent Language Learners</i>	279
Table 51. <i>Assessing the Speaking Skills of Young Language Learners</i>	281
Table 52. <i>Assessing the Speaking Skills of Adolescent Language Learners</i>	282
Table 53. <i>Assessing the Reading Skills of Young Language Learners</i>	282
Table 54. <i>Assessing the Reading Skills of Adolescent Language Learners</i>	284
Table 55. <i>Assessing the Writing Skills of Young Language Learners</i>	286
Table 56. <i>Assessing the Writing Skills of Adolescent Language Learners</i>	287
Table 57. <i>Assessing Grammar: Young Language Learners</i>	289
Table 58. <i>Assessing Vocabulary: Young Language Learners</i>	290
Table 59. <i>Assessing Grammar: Adolescent Language Learners</i>	291
Table 60. <i>Assessing Vocabulary: Adolescent Language Learners</i>	292
Table 61. <i>Scoring the Answers or Performances</i>	294
Table 62. <i>Interpreting and Using the Assessment Results</i>	295
Table 63. <i>Awareness of Ethical Assessment Process</i>	296
Table 64. <i>Ensuring Ethical Assessment</i>	300
Table 65. <i>Reflections on the Improvement of LAL</i>	306

LIST OF FIGURES

<i>Figure 1.</i> Contexts, principles, and practices model.....	28
<i>Figure 2.</i> Differential AL/LAL profiles for four constituencies.....	29
<i>Figure 3.</i> A screenshot of data analysis through MAXDQA Analycis Pro 2020.....	48

LIST OF ABBREVIATIONS

LAL	Language Assessment Literacy
NRA	Norm-referenced Assessment
CRA	Criterion-referenced Assessment
PSET	Pre-service English Teacher
FLPT	Foreign Language Proficiency Test
FLT	Foreign Language Test
TEELT	Testing and Evaluation in English Language Teaching
ELT	English Language Teaching
MoNE	Ministry of National Education



CHAPTER I

INTRODUCTION

Introduction

This chapter involves sections related to the background of the study, problem statement, aim of the research, significance of the study, and theoretical framework.

Background to the Study

Assessment is an indispensable phenomenon for formal education. Assessment is always necessary during the process of teaching and in the end of it whether teaching is done in the form of just presenting information to the learner or not. That is to say, assessment should not always be the *assessment of learning*, which is conducted “to summarize what has been learned” (Harlen, 2007, p. 121). It can also be in the form of *assessment for learning*, which involves the acts to detect the deficiencies in learning and to take the essential steps for the improvement of teaching (Black, Harrison, Lee, Marshall, & Wiliam, 2003, p. 2; Lee & Coniam, 2013; Popham, 2011). In this context, it can be said that teaching, learning, and assessment are the complementary components of a dynamic system (Erickson, 2020, p. 30; Lee & Butler, 2020), “and therefore a teacher’s ability to engage with a range of teaching, learning, and assessment practices is essential” (Berry, Sheehan, & Munro, 2019, p. 113). Teachers, who are one of the critical agents in this system, are to be literate about assessment (Stiggins, 1995) because “assessment literacy is a commodity needed by teachers for their own long-term well-being, and for the educational well-being of their students” (Popham, 2009, p. 11). That’s why, teachers are expected to have enough theoretical and conceptual

knowledge about assessment, skills to prepare the instruments for assessment, and be aware of the principles regarding how the learners can be fairly assessed and the assessment results can be interpreted and used properly (DeLuca & Klinger, 2010). The acquisition process of this literacy by teachers begins when they are trained in the faculty of education in order to be qualified as teachers (Öz & Atay, 2017). Because of this situation, teacher training programs should prepare teacher candidates as assessment literates well enough to facilitate learning and the achievement of the learning objectives (DeLuca & Klinger, 2010). And English language teacher education programs are not exempt from this responsibility. In other words, although language assessment literacy (LAL) has some unique features that differ from those of general assessment literacy (AL) (Giraldo, 2018), it is undeniable that the English teacher candidates should receive sufficient training on language assessment.

Problem Statement

Assessment is one of the most important components of education, and thus, “assessment literacy is central to the quality of education because competencies in assessing student learning lead to informed decisions” (Xu & Brown, 2017, p. 133). For example, every student has to pass his or her grade according to the results of some kinds of assessments. Moreover, the candidates can be accepted or declined by a school with regard to their scores from a high-stakes or proficiency exam. Teachers can also reorganize their lessons by using the means of formative assessment. In addition to these, teachers spend a significant amount of their valuable time on assessment (Stiggins, 1999). Therefore, it is obvious that “all teachers need to possess sufficient assessment literacy so their classroom assessments are at least satisfactory or, preferably, substantially better” (Popham, 2009, p. 7). In other words, they should be “conversant and competent in the principles and practice of language assessment” (Harding & Kremmel, 2016, p. 415). And English teachers are also not exempt from this requirement. That is to say, the English teachers need to have enough knowledge and skills regarding assessment, and they are required to take some criteria into consideration while interpreting and using the assessment results. And these teachers can acquire this knowledge and competence with the help of pre-service teacher education programs because “pre-service teacher education is the central method for preparing competent and confident beginning teachers” (DeLuca & Klinger, 2010, p. 419). However, the related research in the literature indicates that the English teachers do not have adequate literacy for language

testing and assessment (Lee & Butler, 2020; Mede & Atay, 2017; Ölmezer-Öztürk & Aydın, 2018, 2019; Vogt & Tsagari, 2014), although it is a construct “which is central to achieving and maintaining the overall quality of teaching and learning” (Crusan, Plakans, & Gebril, 2016, p. 43). Moreover, the teacher candidates are usually not well-prepared enough to assess their students because they do not take adequate pre-service education to do that (Kavaklı & Arslan, 2019; Ölmezer-Öztürk, 2018; Sarıyıldız, 2018, p. ii). Therefore, it is crystal clear that the development of LAL should be investigated by considering the education given in these departments “because they provide future teachers with their first introduction to assessment and evaluation” (Volante & Fazio, 2007, p. 751) and “assessment education within the pre-service period is necessary to provide a strong foundation for future professional learning” (Deluca & Klinger, 2010, p. 420).

Furthermore, since LAL is a phenomenon the development of which cannot be limited to the effects of only the English language testing and evaluation (ELTE) course, there is a need for examining the process which the English teacher candidates go through during their teacher training and for investigating to what extent the pre-service teacher education enables the teacher candidates to develop their LAL. However, to the best of the researcher’s knowledge, there is not any research study that was conducted to fill this gap.

While some of the previous studies investigated how the ELTE courses (Hatipoğlu, 2015; Sevimel-Şahin, 2019; Şahin, 2019) or workshops (Baker & Riches, 2018) affected teachers’ or teacher trainees’ LAL, some others described the LAL levels of the English as a foreign language instructors (Yastıbaş, 2018) or the pre-service English teachers (Hatipoğlu, 2015). However, they did not focus on the effects of other courses and limited LAL development to the effects of only the ELTE course, whereas LAL is a concept that involves content knowledge in addition to the knowledge related to assessment. That’s why, considering the development of LAL with the possible effects of other courses given in the ELT departments will help the teacher educators and researchers better understand how the pre-service teachers develop the LAL concept during pre-service education. Moreover, this research may present enlightening results for making the necessary changes on the pre-service teacher education program with the aim of improving the LAL level of pre-service teachers.

As another issue to take into consideration, it is acknowledged that LAL development is influenced by the teacher trainees’ experiences and the context in which the teacher candidates exist. However, how LAL development is influenced by these factors has been

investigated by only a few studies (Yan, Zhang, & Fan, 2018). In this regard, there is a need for a study examining the factors influencing pre-service teachers' LAL development process.

Aim of the Study

The present study aims at describing how the pre-service English teachers develop their LAL throughout their training in an English language teaching department. Furthermore, displaying the extent to which the English teacher training program facilitates the acquisition of LAL is another aim of the research. As compatible with these aims, the study will answer the following research questions:

1. What is the junior pre-service English teachers' LAL level?
 - a. What is the junior pre-service English teachers' level concerning the knowledge component of LAL?
 - b. What is the junior pre-service English teachers' level concerning the skills component of LAL?
 - c. What is the junior pre-service English teachers' level concerning the principles component of LAL?
2. What are the expectations of the junior pre-service English teachers for the LAL development process?
3. What are the suggestions of the junior pre-service English teachers for the LAL development process?
4. What is the third grade pre-service English teachers' LAL level?
 - a. What is the third-grade pre-service English teachers' level concerning the knowledge component of LAL?
 - b. What is the third-grade pre-service English teachers' level concerning the skills component of LAL?
 - c. What is the third-grade pre-service English teachers' level concerning the principles component of LAL?
5. What are the reflections of the third-grade pre-service English teachers on the LAL development process?

6. What are the expectations of the third-grade pre-service English teachers for the LAL development process?
7. What are the suggestions of the third-grade pre-service English teachers for the LAL development process?
8. What is the senior pre-service English teachers' LAL level?
 - a. What is the senior pre-service English teachers' level concerning the knowledge component of LAL?
 - b. What is the senior pre-service English teachers' level concerning the skills component of LAL?
 - c. What is the senior pre-service English teachers' level concerning the principles component of LAL?
9. What are the reflections of the senior pre-service English teachers on the LAL development process?
10. How is the LAL development process in the pre-service English teacher training program?

This research does not aim to measure the participants' LAL with the help of a test or a scale. Also, it does not purport to investigate the effects of specifically designed training on improving their LAL levels. Instead, it aims to describe the change in the pre-service teachers' LAL by considering the effects of the training given in their department. Therefore, the study will not reveal any quantitative results about the teacher candidates' LAL level.

Significance of the Study

In the relevant literature, there are few studies carried out on the English teachers' language assessment literacy in the Turkish education context (Mede & Atay, 2017; Ölmezer-Öztürk, 2018; Sevimeş-Şahin, 2019; Şahin, 2019; Yastıbaş, 2018). Furthermore, while some of these studies (Mede & Atay, 2017; Ölmezer-Öztürk, 2018; Yastıbaş, 2018) involve in-service English teachers, the research conducted with pre-service English teachers is very scarce (Sevimeş-Şahin, 2019; Şahin, 2019). Of these studies with the teacher candidates, some investigated the effects of ELTE courses on the development of LAL. On the other hand, some others aimed to determine the prospective English teachers' level of LAL.

How the pre-service English teachers develop their LAL throughout their pre-service education is an uncharted territory of inquiry. Although the assessment courses given in the English language teaching (ELT) departments have significant effects on the development of LAL, taking the other courses into consideration in this research will enable us to look at the issue from a broader perspective. Furthermore, because the current models of LAL are usually theory-driven and deductively developed (Lee & Butler, 2020), the present research drawing on empirical data will provide the background to revise the existing LAL models and will contribute to defining LAL within the pre-service teacher education context.

In addition, “qualitative studies investigating pre-service language teachers’ perspectives about their experience of language assessment training have been few and far between” (Lam, 2015, p. 171). And in consideration of the data collection methods, the existing studies in the literature are generally based on the data from surveys and interviews (Lee & Butler, 2020). However, in this research, some other methods for data collection, such as observations, diaries, interviews, documents, and scenarios, will be employed, and, by this means, the collected data will be triangulated by different sources.

Theoretical Framework

Constructivism and sociocultural theory (SCT) constitute the theoretical basis of this research. In the following sections, some detailed information about the theoretical background of this study will be mentioned.

Constructivism

Constructivism, which emerged as a non-positivist paradigm in the twentieth century and “profoundly influenced the development of qualitative research” (Croker, 2009, p. 6), is a learning and knowledge theory (Fosnot, 2005). Integrating different paradigms related to linguistics, psychology, and sociology (Brown, 2007, p. 12), this theory emphasized the importance of social interaction among people and the learners’ construction of their own reality for the learning process (Brown, 2007, p. 12; Creswell, 2007, p. 21) since “there is no universally agreed upon reality or universal truth” according to the constructivist paradigm (Croker, 2009, p. 6). And this social interaction does imply that the organism is a cognitive subject and that the objects are cognitive structures that are constructed with the

help of previous experiences (Von Glasersfeld, 2005, p. 3). In this regard, learners are seen as agents who are active in the learning process by constructing their own knowledge. Therefore, within the constructivist paradigm, learning “requires self-regulation and the building of conceptual structures through reflection and abstraction” (Von Glasersfeld, 1995, p. 14), and this differentiates constructivism from the behavioristic paradigm, according to which learning occurs as a result of the relationship between stimulus and response (Fosnot & Perry, 2005).

Moreover, Richardson (1997) points out that “constructivism is a descriptive theory of learning (this is the way people learn or develop); it is not a prescriptive theory of learning (this is the way people should learn)” (p. 3). As compatible with this, Von Glasersfeld (1989) states the two principles of constructivism as in the following sentences:

1. Knowledge is not passively received but actively built up by the cognizing subject;
2. The function of cognition is adaptive and serves the organization of the experiential world, not the discovery of ontological reality (p. 114).

Here, ‘adaptation’, which is a concept stemming from biology, is used to refer to the interaction between organisms and the environment they live in (Von Glasersfeld, 2005, p. 3). As compatible with this, Jean Piaget suggested “the idea that what we call knowledge does not and cannot have the purpose of producing representations of an independent reality but instead has an adaptive function” (Von Glasersfeld, 2005, p. 1). In this context, according to the constructivist paradigm, learning, which is a complex and nonlinear process (Fosnot & Perry, 2005), is constructed by the students, and the teachers are responsible for helping the students in this process by creating the opportunities for the construction procedure. Additionally, constructivism focuses on the development of cognition and understanding rather than the change of behaviors or the improvement of skills (Fosnot & Perry, 2005).

Sociocultural Theory

SCT is based on the ideas of Russian psychologist Lev S. Vygotsky. Within this theory, humans are seen as “socially organized entities” (Lantolf, 2007, p. 32). Therefore, interaction, learning with cooperation (Brown, 2007, p. 12), and the effects of the sociocultural environment on learning (Ableeva & Lantolf, 2011) are emphasized in the

SCT. “It is within this social interaction that cultural meanings are shared within the group, and then internalized by the individual” (Richardson, 1997, p. 8). According to the SCT, humans’ cognitive and emotional development is “mediated by others, whether they are immediately present ... or displaced in time and space” (Lantolf, 2007, p. 32). That is to say, Vygotsky (a) proposes a genetic analysis method for development, (b) argues that social life is the source of individuals’ higher mental functioning, and (c) mediational tools and symbols mediate humans’ actions both individually and as a society (Wertsch, 1993, p. 19). This sociocultural interpretation of Vygotsky’s sociohistorical approach “tends to focus on a microlevel analysis of social interaction within the zone of proximal development” (Vadeboncoeur, 1997, p. 25). Vygotsky’s methodology involves four different layers of historical development. The first level, microgenesis, is concerned “with the emergence of particular mental processes (e.g. a particular language ability such as listening comprehension) over relatively short periods of time” (Ableeva & Lantolf, 2011, p. 137). On the other hand, ontogenesis refers to “the subject’s individual developmental history” (Vadeboncoeur, 1997, p. 26). As to phylogenesis, Vygotsky impressed by other scholars’ ideas such as Darwin and Engels (Wertsch, 1993, p. 20), used this term to address the biological evolution of humans. Finally, the last layer is related to the individual’s “sociohistorical development or the development of socially organized activities” (Vadeboncoeur, 1997, p. 26).

This perspective involves such concepts as mediation, zone of proximal development (ZPD), dynamic assessment, and so on. Of these concepts, mediation is very important for SCT, and as regards this, Lantolf (2000) expresses that “the most fundamental concept of sociocultural theory is that the human mind is *mediated*” (p. 1). Lantolf and Thorne (2006) define mediation as “the process through which humans deploy culturally constructed artifacts, concepts, and activities to regulate (i.e. gain control over and transform) the material world or their own and each other’s social and mental activity” (p. 79). Thus, SCT involves the idea that humans prefer to use some cultural artifacts in order to interact with the environment (Lantolf, Thorne, & Poehner, 2015, p. 209). And these can be physical and symbolic (psychological) tools (Lantolf, 2000, p. 1). The physical tools are the objects that enable individuals more easily do the activities in their daily lives by making them physically stronger than they actually are (Ableeva & Lantolf, 2011). On the other hand, the symbolic tools are “auxiliary means to control and reorganize our biologically endowed mental

processes” (Lantolf et al., 2015, p. 210). Although mathematical signs, music, the arts, and charts are some of these symbolic artifacts, language is accepted as the most important of them (Lantolf et al., 2015, p. 210).

Kozulin (2003) also refers to the two faces of mediation as human and symbolic (p. 18). Human mediators can be parents, teachers, peers, or a more knowledgeable person. These people can help the learners while they are engaged in the tasks that can be challenging for them, and thus the human mediators can carry the learners to a higher level through mediation. As to the symbolic tools or mediators, Vygotsky (1978, as cited in Kozulin, 2003, p. 23) cites “casting lots, tying knots, and counting fingers” as the most ancient symbolic tools. In addition to these, as Lantolf et al. (2015, p. 210) mentioned above, writing, maps, “letters, codes, mathematical signs”, and any other tools can lead to cognitive changes in the learners (Kozulin, 2003, p. 26). However, Kozulin (2003) warns us that the symbolic tools should be properly, intentionally, and meaningfully mediated to the learners, and the learners should appropriate and internalize these tools as general instruments in order to use them in other contexts (pp. 24-26).

As for another well-known concept of SCT, it is ZPD, or *obuchenie* in Vygotsky’s Russian terms (Ableeva & Lantolf, 2011). In the field of education, ZPD is generally used to refer to a metaphorical gap between the learner’s actual capacity and the performance of the learner when helped by an adult or a more knowledgeable person. This gap is seen as different for every learner, and therefore, the amount of time every learner needs to fill this gap changes from learner to learner. However, Vygotsky used ZPD with different meanings in different contexts. Kozulin, Gindis, Ageyev, and Miller (2003) explain them as follows:

In the developmental context ZPD is used for explaining the emerging psychological functions of the child. In the applied context ZPD explains the difference between the child’s individual and aided performances, both in situations of assessment and in classroom learning. Finally, ZPD is used as a metaphoric “space” where everyday concepts of the child meet “scientific” concepts provided by teachers or other mediators of learning (p. 3).

As for the dynamic assessment (DA), it is “an approach to understanding individual differences and their implications...[that] embeds intervention within the assessment procedure” (Lidz & Gindis, 2003, p. 99). Focusing on both processes and products related to development (Ebadi & Latif, 2015), DA is an approach to assessment that is based on Vygotsky’s sociocultural theory and especially on the ZPD (Ebadi & Rahimi, 2019; Shrestha & Coffin, 2012). In contrast to traditional types of assessment, DA supports the view that assessment and instruction should not be separated during the assessment process. In other words, DA “offers a monistic view of assessment and instruction” (Ebadi & Rahimi, 2019,

p. 2). Furthermore, there are two different approaches to DA: interventionist and interactionist (Ebadi & Saeedian, 2016; Ebadi & Latif, 2015). While the interventionist approach involves the provision of the same mediation to all learners without taking the learners' individual differences into consideration (Yılmaz-Yakışık, 2012), the interactionist approach supports the provision of mediation during one-on-one sessions, and this mediation can be different for each learner. Therefore, the implementation of DA according to this approach is more effortful and time-consuming than the implementation of DA according to the interventionist approach. However, the provision of one-on-one mediation can be more effective for addressing the learner's ZPD.

Having these theoretical characteristics briefly mentioned above, SCT was chosen as the theory that would be informative for this study because, as Johnson (2009) states, it is possible to “trace teacher learning from a sociocultural perspective by looking at the progressive movement from externally, socially mediated activities to internal mediation controlled by the individual teacher” (p. 17). Additionally, Inbar-Lourie (2008) also suggests the use of the constructivist approach for the development of assessment literacy by stating that “the manner and the process of acquiring assessment literacy [...] assume a constructivist learning approach, whereby the participants and assessment experts form a knowledge community by discussing, critiquing and questioning fundamental issues relevant to their context” (389). Similarly, the pre-service English teacher education context in Turkey involves features compatible with this theory. For example, in this context, there is always interaction between instructors and the teacher candidates as well as among the teacher candidates themselves. Moreover, various physical and symbolic tools such as language, texts, audios, coursebooks, technological tools, and so on are always used intentionally. And it is aimed at carrying the teacher trainees to a more competent and autonomous level by taking their ZPD into consideration. Because of all these reasons, this research will deploy SCT as the theoretical framework.

Definition of the Key Terms

Assessment: It is broadly defined as “the process of collecting information about a given object of interest according to procedures that are systematic and substantively grounded” (Bachman, 2004, p. 7).

Evaluation: It means “making a judgement of the value or worth or meaning of an activity, event, or body of information” (Berry, 2008, p. 7).

Measurement: This term means “the quantitative description of particular characteristics of a class of people, objects, systems, or events” (Berry, 2008, p. 7).

Practicality: It refers to the features of a test which “is not excessively expensive, stays within appropriate time constraints, is relatively easy to administer, and has a scoring/evaluation procedure that is specific and time-efficient” (Brown, 2004, p. 19).

Reliability: It means “the extent to which an assessment produces consistent and stable results” (Green, 2014, p. 239).

Test: It means “a measurement instrument designed to elicit a specific sample of an individual’s behavior” (Bachman, 1990, p. 20).

Validity: “Validity refers to the degree to which evidence and theory support interpretations of test scores entailed by proposed uses of test scores” (AERA, APA, NCME, 1999, p. 9).

Washback: It “refers to the influence of testing on teaching and learning” (Cheng & Curtis, 2004, p. 4).

CHAPTER II

LITERATURE REVIEW

Assessment

Language assessment is an umbrella term that has a significant role in language classrooms (Cheng & Fox, 2017, p. 3) since teachers spend a great deal of their time (DeLuca & Klinger, 2010; Lan & Fan, 2019). It cannot be separated from language teaching and learning. In other words, language assessment is in a mutual relationship with these two components. For example, drawing on the results of language assessments, teachers make inferences about where their students are and where they need to be. Therefore, language teachers play dual roles in a classroom as a teacher and an assessor. Hence, the teachers are informed about the efficacy of their instruction and the steps that should be taken to make the necessary changes.

In the *Standards for Educational and Psychological Testing*, published in 2014, the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (2014) define assessment as “a systematic process to measure or evaluate the characteristics or performance of individuals, programs, or other entities for purposes of drawing inferences” (p. 216). In the literature on language assessment, however, this term has been defined by some scholars as referring to different aspects of it. For example, Purpura (2016) defines assessment as “a systematic procedure for eliciting test and non-test data for the purpose of making inferences or claims about certain language-related characteristics of an individual” (p. 191). Green (2014) also mentions that “language assessment involves obtaining evidence

to inform inferences about a person's language-related knowledge, skills, or abilities" (p. 5). As these definitions imply, the concept of assessment includes various processes such as eliciting information related to the students' competences, scoring or grading the performance or products of the students, making decisions about the students' achievements, and informing teaching and learning. Furthermore, each assessment has some principles which are acknowledged as necessary for criticizing the quality of an assessment. These principles mainly involve reliability, validity, authenticity, practicality, and washback (Brown, 2004, p. 19; Bachman & Palmer, 1996) although some scholars also suggest some qualities with different names such as construct validity, interactiveness, and impact (Bachmann & Palmer, 1996, p. 18).

Reliability

Reliability refers to "the consistency of the scores obtained" (Fraenkel & Wallen, 2009, p. 154). In other words, it means "The extent to which repeated measurements on units (for instance people) yield similar results." (Everitt & Skrondal, 2010, p. 365). As understood from these definitions, reliability is the quality of assessment results rather than the test itself (Bachman, 1990, p. 26; Crocker & Algina, 2008, p. 122). Therefore, it is not accurate to call a test reliable or not. Moreover, reliability is also dependent on the sample of the test takers. The reliability of the scores obtained on a test will change from one group of test takers to another if these two groups have distinctive features.

Some factors affect the degree of reliability, and Brown (2004, pp. 21-22), adapting them from Mousavi (2002, p. 804), lists them as student, rater, test administration, and test. The student can be a source adversely affecting the degree of reliability. To clarify, the issues related to the student's psychology or physiology may cause the interference of more measurement error to the true score. Moreover, test-wiseness and the student's test-taking strategies may also impact the reliability of the measurements. Rater is the other potential factor that can cause inconsistency in the scores because of inattention, bias, subjectivity, inexperience, and fatigue (Brown, 1996, p. 21). In this regard, reliability is divided into inter-rater and intra-rater reliability. The inter-rater reliability means the consistency between or among the scores awarded by at least two raters to a set of the student's performances or products (McNamara, 2000, p. 57; Weir, 2005, p. 34). The intra-rater reliability, on the other

hand, is when a rater gives the same score to the same performance after a while (Douglas, 2010, p. 110; Weir, 2005, p. 34). In order to estimate these two types of rater reliability, correlation is calculated between the two sets of scores (Douglas, 2010, p. 110). In addition, how successfully the test is administered to the students also affects the degree of reliability. In other words, environmental factors such as light, noise, temperature, and comfort of the desk, or factors such as photocopy quality, adequate speaker volume, or cheating, also have the potential to negatively affect the reliability of the measurement results (Brown, 2004, p. 21). The test can also be a source of measurement error that threatens reliability. The length of the test, the number of questions, the ease or difficulty of the questions and the clarity of the instructions affect reliability (Douglas, 2010, p. 105; Black & William, 2006, p. 120).

Reliability has different types, named by estimation methods. These are test-retest, parallel forms, split halves, coefficient alpha, and KR20/21 (Bachman, 1990, pp. 168- 178). The test-retest method involves “consistency of test scores over time” (Popham, 2017, pp. 76-77). In order to estimate reliability with this method, the same test is administered to the same group of students twice within a particular time between the two administrations (Weir, 2005, p. 25). Then, the correlation between the students’ scores on the two administrations is calculated, and the correlation coefficient obtained gives the reliability index. While this is a valuable method, getting the same group together to administer the same test twice is sometimes impossible. Therefore, it may be necessary to administer two supposedly parallel tests to the same group with little time difference. This method called the parallel forms method, requires these two alternative test forms to be as identical as possible regarding test specifications (Weir, 2005, pp. 25-26). The split-half method is another way for the precision of the reliability index. This method “involves administering a test and then dividing the test into two equivalent halves that are scored independently” (Reynolds, Livingston, Willson, 2010, p. 99). Then, the correlation between the scores from the first half of the test and the scores from the second half is calculated by using some formulas for reliability estimation like Spearman-Brown prophecy (Nitko & Brookhart, 2007, p. 73; Reynolds et al., 2010, p. 99). Other reliability estimation methods that can be obtained by applying a test once are Cronbach's alpha and Kuder-Richardson's formula 20 and 21 (KR20/21) (Brown, 2005, p. 179-180). The difference between these approaches to reliability estimation from the split-halves method is that they estimate reliability by calculating the consistency among

responses to all items on the test rather than responses to items on one half of the test (Reynolds et al., 2010, p. 100).

Validity

Validity, the most essential quality for an assessment (Koretz, 2016; Reynolds, Altmann, & Allen, 2021, p. 185), “refers to the degree to which evidence and theory support the interpretations of test scores for uses of tests” (AERA, APA, & NCME, 2014, p. 11). As this definition implies, validity is a property of interpretations made of scores for intended uses, and “what needs to be validated is not the test itself but the use of a test for a particular purpose” (Sireci & Faulkner-Bond, 2015, p. 219). Therefore, it is not correct to express validity as a feature of tests or to talk about the validity of the test (Bademci, 2019, 2022; Kane, 2013). Like reliability, validity also depends on the sample; therefore, the degree of validity may change from one group of test takers to another.

Validity has evolved as a concept and theory in educational and psychological measurement (Bademci, 2019, 2022; Reynolds et al., 2021, p. 191). The *Standards for Educational and Psychological Testing* is usually mentioned as a source to trace this evolution. In the old versions of the *Standards*, validity was divided into three types: content, criterion-related, and construct. The first one refers to the test involving items representative of the content domain that the test intends to measure (Brown, 2005, p. 221). In other words, when an instrument or task, which is a sample selected from the universe of all relevant items or tasks (Weir, 2005, p. 19), is administered to assess students’ knowledge, abilities, or skills, the assessment results should reflect their performance related to the attributes of the target domain. Criterion-related validity involves correlating the results obtained from a test or task with the ones obtained from another test or task assumed to measure the construct in the target (Weir, 2005, pp. 35-36). According to this type of validity, the higher the correlation between the results from two tests, the higher the validity (Brown, 2021, p. 18). Construct validity is the third distinct type of validity. According to Bachman and Palmer (1996), “construct validity pertains to the meaningfulness and appropriateness of the interpretations that we make on the basis of the test scores” (p. 21). Regarding this, Green (2014, p. 81) stated that it would be a disaster for construct validity if the results obtained from a test developed to test whether a person can communicate in daily life using the target language

were used to decide whether that person should go into surgery as a nurse or be hired as an air traffic controller.

However, this “holy trinity” of validity (Guion, 1980) began to be abandoned in the 1970s, and validity began to be conceptualized as a unitary concept. In the 1985 *Standards for Educational and Psychological Testing*, validity was presented as a concept consisting of three interrelated types: content-related validity, criterion-related validity, and construct-related validity (Reynolds, 2021, p. 193). Then the 1999 and 2014 *Standards for Educational and Psychological Testing*, validity is considered as a unitary concept in its most recent form and five sources of validity evidence are presented. These sources are validity evidence based on test content, relations to other variables, response processes, internal structure, and consequences of testing.

Authenticity

Authenticity means “the degree of correspondence of the characteristics of a given language test task to the features of a target language use task” (Bachman & Palmer, 1996, p. 23). As evident in this definition, claiming that a test task is authentic means it can be fulfilled in real life (Brown, 2005, p. 28). Although ensuring authenticity is not simple, Brown (2005, p. 28) states that authentic test tasks have some characteristics. For example, the language used in the test is the same or too similar to the one used in real life. Moreover, the items in that test are given in context rather than in isolation, and the topics appeal to learners’ interests. The items are placed in a meaningful order, like a storyline, rather than just in a sequence. And the most well-known characteristic of authenticity is that the test tasks are very close to real-life tasks. Bachman and Palmer (1996, p. 24) also point out that students’ performance on authentic tests is generalizable to real-life tasks and that this authenticity feature links it with construct validity.

Practicality

Practicality is a desirable quality of assessment tools and practices. Bachman and Palmer (1996) define this term as “the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these

activities” (p. 36). As explained in this definition, the preparation, administration, and scoring phases of the tests or assessment practices should be cost-efficient (Brown, 2005, pp. 28-29). Regarding this, Bachman and Palmer (1996, p. 37) mention three main types of resources: human resources, material resources, and time. Of these, human resources involve the people writing, administering, scoring the test, and giving the clerical and technical support for realizing these steps. As for the material resources, Bachman and Palmer (1996, p. 37) divide them into three types: space, equipment, and materials. The space means the place available for test development and administration. The equipment consists of technological tools such as computers, audio and video recorders, speakers, and printers, while the materials refer to such objects as paper, photos, and documents that will be used for the assessment process. As the last resource, time means the period spanning from the beginning of the test development to the sharing of assessment results and the time spent on each step of the assessment. Moreover, money that will be spent for the use of a test or assessment method should be considered for acknowledging the test or method practical. In consideration of these resources, some types of tests are called more practical in some phases of the test. For example, multiple-choice exams are evaluated as the tools the administration and scoring steps of it are relatively more practical, while preparing them takes too much time for the test maker. On the other hand, alternative assessment methods such as portfolios, projects, and role-plays are acknowledged as less practical for some reasons. The first of them is that these assessment practices take too much time for the students because they are required to produce some products by using the target language. In addition, the teacher also spends too much time and effort rating the products of these practices. Therefore, the teachers and administrators may have some concerns and avoid using these methods, although they are appropriate for communicative assessment of language skills. However, as Bachman and Palmer (1996, p. 37) state, practicality is a concept that should be considered with regard to the context in which the assessment will take place. Thus, alternative assessment methods can also be considered practical under the conditions that the classrooms do not consist of many students, the teachers and the learners have adequate time, and the necessary equipment, materials, and funds are available.

Washback Effect

The washback or backwash effect refers to the effects of testing or assessment on learning and teaching (Cheng & Curtis, 2008, p. 4; McNamara, 2000, p. 73). Washback is a concept that has a bidirectional impact on learning and teaching: positive or negative washback. However, this does not mean that a good test would result in positive washback, while a poor test would absolutely have negative washback on learning and teaching. The reason for that is that there are many other factors in the education system that can be impact the effect of washback (Cheng & Curtis, 2008, p. 11).

In our country, some high school students who are specializing in the area of foreign languages, particularly English, face the negative washback of a nationwide exam. Because this exam is used as one of the gatekeeping elements for admission to university, the students should prepare themselves for this high-stakes exam and get the necessary score to be accepted to an institute of tertiary education. However, while getting prepared for the exam, the students do not focus essentially enough on some skills such as listening, speaking, and writing because these skills are evaluated in the exam. Furthermore, since the teachers feel themselves responsible for preparing their students for the exam, they regulate their teaching accordingly. The teachers may even end up teaching the students some strategies they can use to find the correct answer in this multiple-choice exam.

Formative Assessment

The first introduction of formative assessment dates back to the coinage of the term “*formative evaluation*” by Scriven (1967) in the American Educational Research Association’s (AERA) monograph (Cizek, 2010, p. 5). Scriven (1967) used this term “in connection with curriculum improvement” (Bloom, Hastings, & Madaus, 1971, p. 117). Then, in their influential book, *Handbook on Formative and Summative Evaluation of Student Learning*, Bloom et al. (1971, p. 117) elaborated formative evaluation by regarding it in terms of teaching and learning as well as curriculum construction. In this book, they defined formative evaluation as “the use of systematic evaluation in the process of curriculum construction, teaching, and learning for the purpose of improving any of these three processes” (Bloom et al., 1971, p. 117).

Over the last decades, formative assessment has become more prominent in education, and throughout this process, it has evolved. In addition, it has been called by some different names, like assessment for learning, classroom-based assessment, student-centered assessment, and so on. Although with these different names some aspects were emphasized, all of them involve the basic characteristics that distinguish formative assessment from its cousin, summative assessment (Cizek, 2010, p. 3). Of these, perhaps the most important is that the evidence elicited through formative assessment is to be used by teachers to adjust the instruction in a way that is most suitable for student learning and to be used by students to regulate their own learning as autonomous learners. In a general sense, formative assessment can be defined as the kind of assessment that provides evidence to make adaptations for improving learning. Black and Wiliam (2009, p. 9) state:

Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.

From this definition, Black and Wiliam (2009) refer that for regulating the learning processes “moments of contingency”, which can occur in either a synchronous or an asynchronous way, should be created and focused on during the instruction (p. 10). The synchronous moments of contingency consist of the adaptations made by the teacher at the time of instruction or classroom discussion. In contrast to that, the asynchronous moments of contingency involve the adjustments made for the regulation of subsequent learning. Grades, feedback from the teacher or peers, or the students’ self-assessments following the instruction can be cited among these asynchronous moments.

For conceptualizing formative assessment, Wiliam and Thompson (2017) have drawn on Ramaprasad’s (1983) definition of feedback and the instructional processes relevant to this definition. From this point of view, Wiliam and Thompson (2017) pointed out that in order to acknowledge an assessment as formative, it should have an impact on learning. Therefore, instruction should be adapted in a way that meets the students’ learning needs. In addition to this “big idea” of them, Wiliam and Thompson (2017, p. 127) also suggested the following five key strategies which formative assessment consist of:

1. Clarifying and sharing learning intentions and criteria for success;
2. Engineering effective classroom discussions, questions, and learning tasks;
3. Providing feedback that moves learners forward;
4. Activating students as instructional resources for one another; and
5. Activating students as the owners of their own learning

In the implementation of formative assessment, both traditional classroom tests and other alternative assessment methods such as “observations, oral questioning, class discussions, projects, portfolios, homework, performance assessments, groups work with peer feedback, student self-assessment, and other sources” can be used (Cizek, 2010, p. 7). Regarding how formative assessment can be implemented in class, some different approaches emerged. For example, Popham (2017, p. 291) suggested a four-level approach. According to it, the first level refers to the teachers’ adjustments related to how they teach. These adjustments are made by the teachers, drawing on the evidence elicited via assessments, if there is a need for them to enhance the quality of instruction. As for the second level, it involves the students’ self-regulation on their learning strategies. The students can decide to make changes on the cognitive strategies they use while trying to learn any content or improve their skills. A combination of the first and second levels creates a shift in the classroom climate from teacher-dominated and grade-oriented to student-based and learning-oriented. Finally, the fourth level of formative assessment can be implemented to expand the implementation of formative assessment to the whole institution or district with the help of in-service training or communities of practice.

Furthermore, Wiliam and Black (1998) also suggested a five strategies approach for explaining how formative assessment can be carried out in the classroom. These strategies can be ordered from the most teacher-oriented to the most student-oriented. The first one of them is related to the clarification and sharing of the learning goals and success criteria by teachers and to their understanding by students. The thinking behind this strategy is that the learners will be more successful in accomplishing the learning goals if they have any idea about the reason why they are trying to learn. The second strategy is about the methods that can be exploited for obtaining evidence regarding students’ learning. Related to this strategy, the creation of a discussion environment in the classroom and the usage of productive activities and tasks can be useful for making decisions about the necessity of the adjustments that will be made to the teachers’ instruction and students’ learning. The provision of teacher feedback constitutes the third strategy. This kind of feedback should guide the students about how they can be more successful in their upcoming performances. As another type of feedback, peer feedback can also be employed in language classrooms. The deployment of this fourth strategy can be beneficial for both students. However, since the students might misinform one another about the issue at hand, the teacher should exercise caution when

considering peer feedback. Therefore, the presentation of some guidelines that are to be followed during the peer feedback sessions can be suggested. The last strategy covers self-assessment and self-regulation of learning strategies. The accomplishment of this strategy aimed at making the learners autonomous agents in the learning process can take a long time for the students who have had fewer experiences in their educational lives. As can be understood from these approaches, feedback takes a significant place in formative assessment. The sources of feedback can be teachers, peers, the student's self, or technology (Andrade & Brookhart, 2016, p. 298).

Panadero, Jonsson, and Strijbos (2016, p. 311) mention how some instruments of formative assessment can be used to facilitate the students' engagement in the assessment. For example, portfolios can be one of these tools. Portfolio is a term used to refer to the collection of pieces of work produced by students. The aim of the adoption of portfolios is to gather evidence regarding the development of the learner's performance over time.

Self-Assessment

Self-assessment refers to the assessment process in which students assess their own products or performances in light of the predetermined learning goals and criteria and take some actions in order to improve their products or performances. In other words, self-assessment is "the evaluation or judgment of 'the worth' of one's performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes" (Klenowski, 1995, p. 146). As Black et al. (2004) emphasized, "self-assessment is essential to learning" (p. 14).

Self-assessment can occur in a formative or summative way (Boud & Falchikov, 1989; Ross, 2006). If self-assessment is made for formative purposes, it can be used to make adaptations for the improvement of learning. On the other hand, if self-assessment is used with summative intentions, it can be conducted to detect whether the students have learned or for the assignment of grades. However, Andrade and Valtcheva (2009) argue that the self-assessment should only be used for formative purposes rather than student deciding on his or her own grade. With this perspective in mind, they call the procedure of students engaging in determining their own grades self-evaluation (Andrade & Valtcheva, 2009).

Andrade and Valtcheva (2009) pointed out that the self-assessment process consists of three main steps: the articulation of expectations, the application of self-assessment, and the

revision of the performance or work. Similarly, Wylie and Lyon (2020, p. 172) refer to three steps for the self-assessment process by mentioning the steps suggested by Bailey and Heritage (2018) and McMillan and Hearn (2008). The first of these steps is to ensure that the students understand what is expected of them and the criteria against which their performance will be compared. To enable the students to understand these expectations and the criteria, the teachers should provide the learners with adequate information in order to raise their awareness about them before their performance. Otherwise, the students may avoid taking more responsibility for completing the tasks or performing in the expected way. In the second step, the learners compare their performance or products against the desired performance or criteria. However, in order for the students to make these comparisons, they should not only know and understand the learning expectations or criteria but also have enough practice comparing their work to the criteria. The third step involves the students taking some actions with the aim of improving their performance or products by drawing on the results of their self-assessment.

Ross (2006) also referred to four dimensions concerning the training of students about self-assessment. As the first dimension, it was suggested that “the rubric uses language intelligible to students, addresses competencies that are familiar to students, and includes performance features they perceive to be important” in order to obtain higher reliability and validity in the self-assessment process (Ross, 2006, p. 8). Secondly, the students should be taught how they can implement these criteria while assessing their work or performance. For this instruction, Ross (2006) suggests “teacher explanations of each criterion, teacher modeling of criteria application, and student practice in applying the rubric to examples of student work (including their own)” (p. 9). The third step is the provision of teacher feedback on self-assessment, with the justification that this feedback will ensure the triangulation between the teacher-implemented assessment and the peer assessment. Lastly, it is emphasized that the use of the evidence from self-assessment for the enhancement of student performance can be facilitated with the help of the performance tracking forms, “games, conferences, and menus of examples to support goal setting” (Ross, 2006, p. 9). The quality of self-assessment can be improved with some practice. For example, if the learners have enough experience in peer-assessment, they can also have a better chance to enhance their performance in self-assessment (Topping, 2018, p. 16).

Peer Assessment

Peer assessment is another type of assessment that helps the students in the learning process and is “an important complement to self-assessment” (Black et al., 2004, p. 14). Topping (2018, p. 1) defines peer-assessment as

an arrangement for learners to consider and specify the level, value, or quality of a product or a performance of other equal-status learners, then learn further by giving elaborated feedback to and discussing their appraisals with those who were assessed to achieve a negotiated agreed outcome.

Peer assessment can be deployed in classrooms in the form of distinct variations. Concerning their functions, however, they can be classified under three major categories: summative, formative, and both formative and summative peer assessment (Li, Gao, & Guo, 2019). Summative peer assessment encompasses the assignment of grades or scores to the products or behaviors of a peer by another peer who acts as an assessor. On the other hand, formative peer assessment can be used for the improvement of a piece of work and involves useful feedback for its development. Therefore, the assesseees have the opportunity to revise their works in order to bring them to a higher quality level.

Peer feedback provided within the peer assessment is more available to the learners than teacher feedback. Additionally, peer assessment enables the students to gain critical insights about their own work while making judgments about others’ products (Topping, 1998). Thus, the students’ awareness about which strategies are useful for their learning will increase which can be beneficial for the development of student autonomy.

However, there are also some concerns related to the benefits of peer assessment for learning. Of these concerns, for example, the correctness of the peer assessment might be less than that of the teacher assessment (Topping, 2018, p. 18), especially when the criteria for assessment are not shared with the students. Additionally, the peers may have difficulty criticizing the work of their peers, particularly in terms of the subject content, and some problems with the objectiveness may arise.

Summative Assessment

Summative assessment is the kind of assessment that generally first comes to people’s minds when they hear the words “assessment”, “testing”, or “measurement”. Summative assessment has been very frequently used by teachers, and some significant decisions that can affect the educational lives of students have been made by drawing on the results of

summative assessments. For instance, the decisions on whether the students will pass to another grade or not, whether the scholarship of the students will continue or not, or whether the student is accepted to a program or not are made with the help of the summative assessments. Although the importance of these decisions can change, they have the potential to affect the students' learner identity. Furthermore, it is a fact that the summative assessment results attract the attention of most parents.

Having such an important place in education, summative assessment "refers to the processes and instruments that provide a general and final assessment of student's learning within a given course or learning unit" (Coombe, 2018, p. 39). Likewise, Bloom et al. (1971) also emphasize that "summative evaluation is directed toward a much more general assessment of the degree to which the larger outcomes have been attained over the entire course or some substantial part of it" (p. 61). For labeling an assessment as a summative assessment, Cizek (2010, p. 3) notes that it should have these two main characteristics: (1) the assessment should be administered at the end of an instructional period such as a unit, semester, or year; (2) categorizing a student's or system's performance should be the primary aim of the assessment.

When the content of the ELTE courses is taken into consideration, it is obvious that more importance is given to summative assessment in these courses. For example, in her study, where she examined 36 ELTE syllabi in total and conducted interviews with the ELTE instructors, Şahin (2019) found that "ELTEC in Turkey reflects the features of summative assessment rather than formative assessment" (pp. iv-v). Similarly, Sevimeş-Şahin's (2019) research also revealed that the ELTE course instructors and in-service and pre-service teachers "concentrated on the testing notion, summative evaluation, and only the purposes of diagnosis and achievement in language testing" without much mentioning "formative assessment or other types of testing and assessment" (p. 174). However, summative assessment may have some "unintended, often destructive, consequences for both learning and motivation" (Andrade & Brookhart, 2016, p. 295).

Norm-referenced and Criterion-referenced Assessment

Norm-referenced (NRA) and criterion-referenced assessment (CRA) are two major approaches to assessing the performance of students. Of these, the NRA, which is the dominant assessment approach of the past centuries (Brown & Hudson, 2002, p. 1), entails

making comparisons between the score of each assessee and the mean of the total group scores. That is to say, each learner's achievement or performance becomes meaningful in light of the achievement or performance of the norm group. The NRA assumes that the test takers' scores distribute along a bell-shaped curve, the left and right edges of which include the less frequent scores in the group (Douglas, 2010, pp. 67-68). Regarding the advantages of the NRA, McNamara (2000) states that the NRA "allows for distinct levels of performance to be defined, and allows for distinctions between individual performances to be made" (p. 63). In contrast to these advantages, the NRA inherently includes the disadvantage of creating competition among the students regardless of every student's pace of learning. Additionally, the NRA is criticized for the risk of revealing unreliable and invalid results, especially when the sample is small (Morgan, Dunn, Parry, & O'Reilly, 2004, p. 24).

On the other hand, the CRA, which was first introduced and popularized by Glaser (1963) and Popham and Husek (1969), is based on the idea that the performance or work of the examinee should be compared with a particular criterion or a score predetermined by the examiners or assessors. Although the criterion can vary according to the purpose of the test, the CRA includes a criterion that the test takers should meet at least in order to be successful (Douglas, 2010, p. 68). For example, the examinees may have to score 90 out of 100 on an exam that is administered as the preliminary stage for acceptance to a program. In the CRA, since the learners' success is decided by the comparison of their performance with the specific criteria, it can be motivating for the learners to do their best (McNamara, 2000, p. 64).

Brown (2005) explains the differences between the NRA and the CRA in terms of these six characteristics: interpretation, measurement, purpose, score distribution, structure of the test, and knowledge about test content. Regarding the interpretation, the NRA involves comparing each student's performance with the performance of other students who took the same test. In the CRA, however, a comparison is made between the student's performance and the part to be learned or the objectives to be achieved. While the NRA allows the general measurement of language abilities or proficiency, the CRA is appropriate to measure specific language-related points or abilities.

Diagnostic Assessment

Language teachers need to detect their students' strengths and weaknesses in language learning in order to adapt the instruction and improve the students' deficits (Brown, 2005). Therefore, diagnostic assessment emerges as a necessary way of obtaining the required knowledge for doing that. Defined in various ways by the scholars, the diagnostic assessment has some unique features that help us differentiate it from other types of assessment (Lee, 2015). Firstly, it allows the assessor to get information about not only the strengths but also the weaknesses of the students' knowledge, skills, and abilities related to the target language by focusing on their past learning experiences. Secondly, a set of feedback about the results of the assessment is provided to the stakeholders, such as the learners, teachers, administrators, or parents. Thirdly, remedial activities or instruction are conducted to improve the learners' weaknesses and help them attain the desired level of proficiency. Feedback is a necessary element of diagnostic assessment, and various types of it need to be provided according to the needs of the learners. For example, they can be hints, clues, or computer-based feedback. While the learners can get the feedback after the assessment, they can also receive it during the assessment process, immediately after their response to an item in the computer-based assessments.

Performance-based Assessment

The notion of communicative language teaching led to a shift in the field of language assessment from traditional or structured response testing to alternative or constructed response assessment in the 1970s (Savignon, 1991). As a result, performance-based assessment or performance assessment has become more prominent, and it has been called by different names like alternative, authentic, or task-based performance assessment. Although there is not a widely acknowledged definition of performance-based assessment, it can be defined as a kind of assessment involving the direct observation of language skills through authentic or simulated tasks (Brown, 2005). In contrast to traditional assessment, performance-based assessment allows the assessor to observe the learner's competence, which is elicited through authentic or simulated tasks. According to the focus of assessment, McNamara (2000, p. 121) mentioned strong and weak versions of performance-based assessment. In the strong version, the accomplishment of the task is important, and the target

language is used for doing it. On the other hand, the weak version places more emphasis on the language used during the task completion process than on whether or to what extent the task is accomplished. Written and oral products, interactions, and individual or group performances are the products produced within the performance-based assessment.

Dynamic Assessment

Dynamic assessment (DA) is an approach to formative assessment (Shrestha, 2012, p. 3), and it differs from conventional assessment in which the learner's stable performance or ability is assessed through various assessment practices. However, DA, which is based on Vygotsky's sociocultural theory of mind, involves the assessor's feedback to the learner during the assessment. Zone of Proximal Development (ZPD) is an important concept in the DA framework. The ZPD is used to refer to "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86).

Furthermore, there are two main approaches to DA: interventionist and interactionist. The interventionist approach involves mediation provided according to predetermined criteria. On the other hand, the interactionist approach involves the mediator providing mediation without paying attention to any predetermined and specific criteria or structure. As can be understood from this difference between the two DA approaches, the first one is more appropriate for the DA conducted with a group of learners, while the latter is more suitable for the one-to-one DA sessions between the mediator and the learner.

Assessment Literacy

The American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) (1990) set seven standards for the development of teachers in terms of "assessment, covering skills and principles in assessment development, administration, scoring, and use" (Yan, et al., 2018, p. 159). These standards that are accepted as the first step of defining the competencies that the teachers should be skillful about were stated in the Standards (AFT, NCME, & NEA, 1990, p. 1) as below:

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

Then Stiggins (1991) coined the term *assessment literacy* (AL) in his article published in *Phi Delta Kappan* and defined it by explaining the characteristics of assessment literate and illiterate people. For example, Stiggins (1991) noted that assessment literates “have a basic understanding of the meaning of high- and low-quality assessment and are able to apply that knowledge to various measures of student achievement” (p. 535). On the other hand, he pointed out that “assessment illiterates don’t understand what it takes to produce high quality achievement data and so do not evaluate critically the data they use” (p. 535). After this coinage of the term “seen [...] as a sine qua non for today’s competent educator” (Popham, 2009, p. 4) or “critical for teachers” (Crusan, Plakans, & Gebril, 2016, p. 43), assessment literacy has been diversely defined. For example, Mertler and Campbell (2005) defined assessment literacy as “knowledge of and abilities to apply assessment concepts and techniques to inform decision-making and guide practice” (pp.15-16). More recently, Crusan et al. (2016) remarked that “assessment literacy is not just about content or delivery but how this content is enmeshed with teachers’ knowledge, beliefs, and practices” (p. 45).

As mentioned above, with regard to the features of assessment literate teachers Stiggins (1995) stated that they can differentiate between sound and unsound assessments and that the intricate theoretical side of assessment cannot intimidate them. Inbar-Lourie (2008) also emphasized that being assessment literate “means having 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” (p. 389). Similarly, Gottheiner and Siegel (2012) also point out that “a teacher with assessment literacy can interpret assessment data, communicate with students about the status of their understanding, and employ this information to set short- and long-term learning goals” (p. 533). As seen from these definitions and characteristics above, AL generally refers to three domains: knowledge, skills, and principles, although “what might be the exact components of assessment literacy is still under discussion” (Lan & Fan, 2019, p. 114).

Language Assessment Literacy

LAL, the idea of which was first suggested by Brindley (2001, p. 126), is a different concept than the AL, although it “shares components with assessment literacy” (Lan & Fan, 2019, p. 114). Berry et al. (2019) state that the experts engaged in language assessment associated the AL with the field of language teaching. And Inbar-Lourie (2013) stresses that the nature of the LAL concept has not been exactly defined, and there are uncertainties about which components are indispensable, less essential, or ignorable. That’s why some scholars have defined the LAL concept, and some others also proposed various theoretical models for describing its components.

For example, Vogt and Tsagari (2014), naming the concept ‘language testing and assessment’ (LTA) literacy, defined it “as 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). Considering the difference among the other stakeholders in education, Taylor (2009) defined LAL as “the level of knowledge, skills, and understanding of assessment principles and practice that is increasingly required by other test stakeholder groups, depending on their needs and context” (p. 24). Lam’s (2015) definition is like “teachers’ understandings and mastery of assessment concepts, measurement knowledge, test construction skills, principles about test impact, and assessment procedures which can influence significant educational decisions within a wider social context” (p. 172). In contrast to these non-empirical definitions, Fulcher (2012) provided the following comprehensive definition for AL by drawing on the findings obtained through the survey:

The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom based tests, familiarity with test processes, 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 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).

As for the models concerning LAL, Davies (2008) proposed two trends for AL by examining the textbooks in the field of English language testing. The first trend is a skills + knowledge approach. And the textbooks organized according to this approach focused on the theoretical knowledge about language assessment and the skills related to item writing, statistical analysis, and computer programs for testing. On the other hand, the second trend refers to a

movement from the skills + knowledge trend to skills + knowledge + principles. As different from the first one and accepted as the “current attempt”, this trend involves ‘principles’ that “concern the proper use of language tests, their fairness and impact, including questions of ethics and professionalism” (Davies, 2008, p. 335).

As an illustrative form of his working definition, Fulcher (2012) also provided a model that consists of three layers: contexts, principles, and practices. As Fulcher (2012) noted, the need for the components at each layer may be different for each stakeholder. However, he did not make any explanations about which stakeholders may need to rely on which components in this model. Therefore, it can be stated that although Fulcher (2012) provided the most comprehensive definition for AL/LAL, there is uncertainty about its interpretation for different stakeholders.

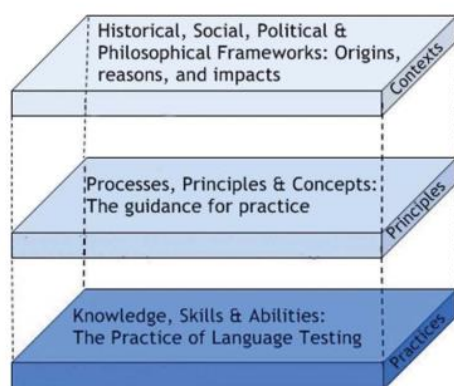


Figure 1. Contexts, principles, & practices model. Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.

Drawing on Bybee’s (1997) categorization and Kaiser and Willander’s (2005) expansion of this categorization, Pill and Harding (2013) suggested five stages for the development of LAL. As given below, these include illiteracy, nominal literacy, functional literacy, procedural and conceptual literacy, and multidimensional literacy stages (Pill & Harding, 2013, p. 383).

Illiteracy: Ignorance of language assessment concepts and methods

Nominal literacy: Understanding that a specific term relates to assessment, but may indicate a misconception

Functional literacy: Sound understanding of basic terms and concepts

Procedural and conceptual literacy: Understanding central concepts of the field, and using knowledge in practice

Multidimensional literacy: Knowledge extending beyond ordinary concepts including philosophical, historical and social dimensions of assessment

In the special issue of *Language Testing* dedicated to LAL, Taylor (2013) also proposed profiles about how LAL develops for four different stakeholders: (a) test writers, (b)

classroom teachers, (c) university administrators, (d) professional language testers as shown in the following figure. As depicted in this figure, Taylor (2013) assumes that LAL has eight components, the importance of which is shown for each stakeholder with numbers from 0 to 4. For example, while language pedagogy has the utmost importance for a language teacher, it is the least important component for a university administrator. On the other hand, knowledge of theory is moderately necessary for classroom language teachers, while test writers and professional language testers need this knowledge more.

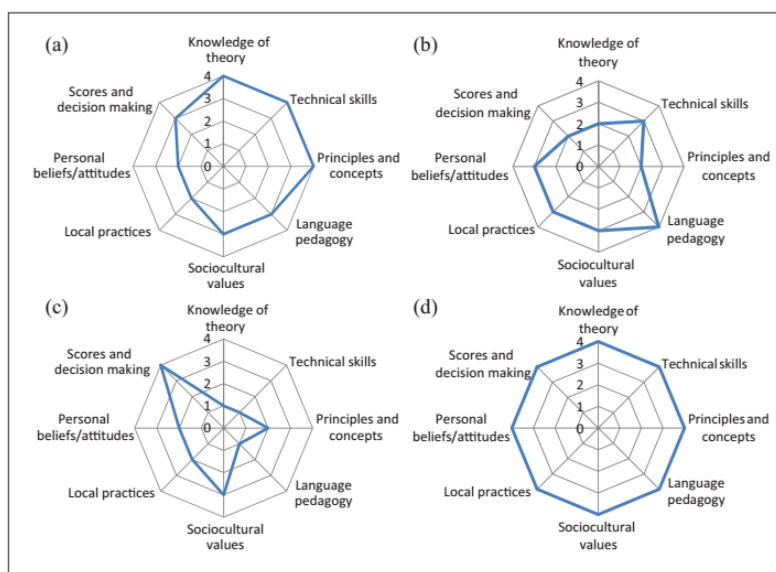


Figure 2. Differential AL/LAL profiles for four constituencies. Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language testing*, 30(3), 403-412.

Although the models mentioned above suggest various components for LAL, all of them acknowledge that LAL is a multifaceted concept (Lee & Butler, 2020). And as mentioned for AL, LAL literature has also referred to these three general components while conceptually defining LAL: theoretical knowledge, language assessment skills, and ethical principles (Baker & Riches, 2018).

Related Studies Conducted in Other Countries

In 2013, *Language Testing*, which is an academic journal focusing on language testing and assessment, dedicated its special issue to LAL. In addition, there are several studies carried out on LAL. Of these investigations, Jeong (2013) investigated whether the instructors' backgrounds are important on the content and structure of the language assessment courses.

To this end, she conducted a mixed-methods research study in which the language testers, whose research interests involved language assessment, and the non-language testers, who were interested in other topics of research related to the field of applied linguistics, participated. While a survey was adopted to collect the quantitative data, semi-structured interviews were conducted to elicit the qualitative data. The findings of this research indicated the importance of the instructors having adequate level of LAL for determining the content of the language courses since the language testers were more confident in teaching technical skills of assessment.

Vogt and Tsagari (2014) conducted a mixed-methods study to investigate the LAL levels and training needs of teachers in seven European countries, including Turkey. The researchers collected the quantitative data through a questionnaire and the qualitative data through semi-structured interviews held with teachers from Cyprus, Germany, and Greece. According to the findings, the study revealed that the teachers did not think their LTA literacy was developed enough and that they did not receive the necessary pre-service training related to this area.

In his qualitative research, Lam (2015) surveyed “the language assessment training backdrop among Hong Kong teacher education institutions (‘TEI’), and specifically investigates whether and how two language assessment courses facilitate and/or inhibit pre-service teachers’ development of LAL” (p. 170). Lam (2015) used documents and websites pertaining to the teacher education programs, focused group and individual interviews, and course materials to collect data. This research indicated that five TEIs provide the teacher candidates with basic language assessment training, although there are some limitations in terms of the provision of core assessment courses to all pre-service language teachers. Additionally, the study revealed that the pre-service teachers, stating that the language assessment training given in TEIs is too academic and inefficient for the cooperation of theory and practice, did not perceive the training as adequate enough to prepare them to be language assessment literate teachers.

Tsagari and Vogt (2017) investigated the perceptions of the teachers’ LAL from Greece, Cyprus, and Germany. To this end, they collected qualitative data through semi-structured interviews. Their participants were foreign language teachers who did not work as item writers or examiners of standardized tests, only as regular EFL teachers conducting classroom assessments. The findings of this research revealed that the teachers’ perceived

level of LAL was generally low, and they made use of pen-and-paper tests which involved classical testing items such as sentence completion, grammar- and vocabulary-oriented questions, and writing the meanings of the given words in their mother tongue. Moreover, the research found that these teachers had a tendency to exploit traditional tools of assessment, and most of them (Cypriot and Greek teachers) adopted transparent criteria to score their students' answers while some of them (German teachers) relied on their intuitions to do so. Consequently, all these findings indicated that they had not taken adequate pre-service training on language testing and assessment.

Baker and Riches (2018) also conducted a mixed-methods research study with the aim of examining how Haitian English teachers and language assessment experts developed their LAL during the workshops held with language assessment specialists. In addition, the researchers investigated the challenges that the teachers face and their successes during the workshops. To these ends, they drew on the teachers' comments about the revised national exams during the workshops, an open-ended survey, and semi-structured interviews. The findings of this research displayed that the teachers developed their knowledge, skills, and ethical practices related to language assessment, while the facilitators of the workshops improved their propositional knowledge regarding the local context. Additionally, the teachers reported they had difficulty making decisions and coming to a consensus with their peers, while they expressed their appreciation for collaboration.

Giraldo and Murcia (2018) conducted an action research study, which aimed to investigate the effects of an assessment course for the pre-service English teachers in Colombia, and presented preliminary findings from this research. The participants of the study consisted of student teachers, professors and one language teacher training expert. To collect data, they used a questionnaire, a survey with open-ended questions, an interview, and the researchers' journal entries.

Firoozi, Razavipour, and Ahmadi (2019) investigated Iranian EFL teachers' needs related to language assessment subsequent to the assessment reform. In order to collect qualitative data, the researchers conducted in-depth interviews and examined the documents regarding the reform. Following inductive analysis of the data, the findings showed that the teachers need training on theoretical knowledge of assessment and on assessing productive and higher-order thinking skills.

More recently, Roslan, Nishio, and Jawawi (2022) investigated 56 Bruneian and Japanese English language teacher candidates' LAL. For collecting data, they administered a questionnaire to all participants and conducted interviews with fourteen of them. This study found that the teacher candidates did not have enough knowledge of LAL.

Related Studies Conducted in Turkey

Hatipoğlu (2015) investigated the pre-service English teachers' expectations from the ELTE course given at Middle East Technical University (METU). In her research, Hatipoğlu (2015) employed a survey for needs analysis and held semi-structured focus group interviews. And she found that the pre-service teachers' assessment knowledge was limited and that the participants expected to take training about test construction, measurement, and preparing their students for different kinds of exams.

In their study, Mede and Atay (2017) aimed to investigate the university instructors' language testing and assessment literacy. To this end, they conducted a mixed-methods research study and collected data through an online questionnaire and focus group interviews. The findings of the research indicated that the language testing and assessment literacy of the Turkish instructors of English is at a low level.

Furthermore, there are some studies conducted as doctoral dissertations. For instance, Yastıbaş (2018) conducted a multiple-case study in order to research how the Turkish EFL teachers working at a Turkish higher education institution implement LAL in their classrooms. To collect data, the researcher used various methods: "individual interviews, think-aloud protocol, observations, focus group discussion and document analysis" (Yastıbaş, 2018, p. v). Content analysis was conducted to analyze the data from these sources, and various findings were obtained. For example, it was found that the teachers were literate enough in terms of language assessment, and they acquired this literacy by assessing and evaluating their students. Furthermore, the instructors stated that for assessing their students, they generally used the materials in course books and on the internet because preparing their own materials may be labor-intensive and take too much time.

Ölmezer-Öztürk (2018) aimed to develop a Language Assessment Knowledge Scale (LAKS) and measure the level of language assessment knowledge of the Turkish EFL instructors working at preparatory schools of Turkish universities. With these aims in mind, initially she prepared a scale involving 60 items measuring four constructs and then

administered this scale to the instructors online. As a follow-up to the findings from the scale, the researcher collected qualitative data through the open-ended questions sent to some of the participants via email. The results of the scale showed that the instructors did not have enough knowledge about language assessment in general, and the findings from the open-ended questions supported these results. More specifically, the participants had the highest mean score in reading while their mean score in listening was the lowest. As for the effects of the participants' demographic features on their LAK level, the research found that only working at a testing office or not was significantly effective. Furthermore, the instructors reported that the pre-service and in-service education were not sufficient enough, and therefore, they expressed their need for workshops and trainings about language assessment. In her doctoral dissertation, Şahin (2019) investigated the ELTE course instructors' planning of the ELTE courses, their assessments and the difficulties they face in these courses, and the pre-service English teachers' evaluations regarding their LAL training that they receive during the ELTE courses. As data collection materials, the researcher used semi-structured interviews, syllabi of ELTE courses, and a LAL training questionnaire. This mixed-methods research revealed that the ELTE course, focusing more on the summative and theoretical sides of assessment than the formative and practical sides of it, is not sufficient to develop the pre-service English teachers' LAL. In addition to this, the pre-service English teachers stated that they received enough knowledge about LAL while they emphasized that they did not take enough training on "alternative assessment tools, formative assessment, and assessing productive skills" (Şahin, 2019, p. v).

In Sevimel-Şahin's (2019) descriptive research, the aim was to detect the pre-service English teachers' LAL levels and their needs for assessment training in the ELTE course. Conducted with course lecturers from different universities, novice EFL teachers, and pre-service EFL teachers taking the ELTE course, the research involved open-ended questions and documents about the course content. Then the data from the open-ended questions was analyzed through content analysis while document analysis was conducted to analyze the data from the webpages of the ELTE course given at the selected universities. The findings indicated that although the stakeholders thought that the course was beneficial for learning the fundamental theoretical knowledge about language testing, it had some limitations as well. For example, the alumni and the course-takers stated that they did not have any chance

to practice in the course, and some topics regarding formative assessment, statistics, assessment of integrated skills, and ethical principles were not covered enough.

As a conclusion, when these studies mentioned above are examined, it is possible to say that they generally focus on one of these topics: LAL levels and needs for training of language teachers or the effects of the ELTE course on the LAL levels of pre-service teachers. Based upon this conclusion, it can be said that there is a need for a study investigating how the LAL concept emerges and develops during pre-service teacher education. And this research aims to fill this gap in the LAL literature.



CHAPTER III

METHODOLOGY

Research Design

This study has a qualitative research design. As the most specific feature of it, this design “entails collecting primarily textual data and examining it using interpretive analysis” (Croker, 2009, p. 5). Requiring a relatively longitudinal data collection and analysis process, qualitative research enables the researchers to get deeper insights about the research topic in question (Croker, 2009, p. 8). In addition, this kind of research is acknowledged as more appropriate for unfolding the development of a concept. And the use of qualitative research is beneficial for refining and developing the previously suggested conceptual framework through empirical data (Chaudron, 2000).

Qualitative research has been the focus of some different approaches, such as case study, ethnography, grounded theory, phenomenology, and narrative research (Creswell, 2007, p. 35; Duff, 2008, p. 38). Of these, the case study is the approach that guides this research. Being widely adopted in qualitative research studies, the case study was defined in various ways by the methodologists. For example, Merriam (1988, p. 16) described it with the following sentences:

The qualitative case study can be defined as an intensive, holistic description and analysis of a single entity, phenomenon, or social unit. Case studies are particularistic, descriptive, and heuristic and rely heavily on inductive reasoning in handling multiple data sources.

More recently, Duff (2020, p. 145) explained the case study as a method that “offers strong heuristic properties as well as analytic possibilities for illustrating a phenomenon in very vivid, detailed, and highly contextualized ways from different perspectives”. Furthermore, Yin (2003, p. 3) classified the case study into three types by taking their purposes into

consideration: exploratory, descriptive, or explanatory. An exploratory case study of these aims to guide the subsequent study in terms of the research questions and hypotheses that will be investigated or give information about how feasible the research procedures are. On the other hand, the aim of a descriptive case study is to portray a situation or phenomenon in a particular context. As for the explanatory case study, its purpose is to present information about how the events occurred in a context by focusing on the cause-and-effect relationships. In addition, case studies are also labeled as single- and multiple-case studies according to the number of cases they include. Single-case studies, which are also called classic case studies (Yin, 2003, p. 46), are conducted to investigate only one case, which can consist of an individual, a pair, a group, or a community (Duff, 2020, p. 147). In multiple-case studies, in contrast, the investigators focus on at least two cases (Fraenkel, Wallen, & Hyun, 2023, p. 390), and these cases can be individuals diverging in some aspects, different classrooms, schools, or programs.

These two kinds of research designs have their own strengths and weaknesses. For example, the single-case design is considered less robust than the multiple-case design because the findings obtained from one case are not considered compelling enough (Fraenkel et al., 2023, p. 390). Also, the single-case design involves the risk of misrepresentation because “a case may later turn out not to be the case it was thought to be at the outset” (Yin, 2018, p. 99). Additionally, conducting research based on the multiple-case design requires much more time, effort, and resources (Fraenkel et al., 2023, p. 390; Yin, 2018, p. 102). In consideration of these types of case studies, the present research is a descriptive multiple-case study which aims to describe how the pre-service English teachers develop their LAL.

Participants

The participants of this research were conveniently selected. That is to say, convenience sampling, involving the selection of “a group of individuals who (conveniently) are available for study”, is the strategy that is used in selecting the participants for this research (Fraenkel et al., 2023, p. 99). In this vein, to reach the participants, the researcher sent messages about the research to the classroom groups on WhatsApp and contacted some students, asking them to invite their friends to participate in the study. Moreover, the researcher contacted the academicians to inform the students about the research. As a result of these efforts to find participants, twenty students studying at different grades in the English Language Teaching

Department of Gazi University participated in the study. Of these participants, eighteen were from Turkey, and their native language was Turkish while two (P1-3 and P4-2) were from Turkmenistan, and their native language was Turkmen. Although Turkish was not their native language, the participants preferred to answer the questions in Turkish rather than in English and did not have significant challenges in understanding and speaking Turkish.

Furthermore, in this study, it is assumed that the participants will be “regular English teachers” after they graduate from their department, as in Vogt and Tsagari’s (2014) research. That is to say, they will teach English to young learners and teenagers in public or private schools, and they do not prepare items for any standardized exams or work as examiners for any testing offices after graduation. For this reason, the present research appeals to only individuals who plan to teach English in any school from primary to upper secondary school level of education.

Table 1

Demographic Information of the Participants

Grade	Participant	Gender	Age	Graduation	Prep-class in High School	Prep-class at University	GPA
1	P1-1	Male	19	Anatolian High School	NO	NO	1,58
	P1-2	Male	19	Anatolian High School	NO	NO	3,50
	P1-3	Female	20	A High School in Turkmenistan	YES	YES	2,97
	P1-4	Female	20	Open Education High School	YES	NO	3,30
2	P2-1	Male	19	Anatolian High School	NO	NO	2,98
	P2-2	Male	20	Social Sciences High School	YES	NO	3,74
	P2-3	Female	20	Anatolian High School	NO	NO	3,35
	P2-4	Female	21	Anatolian High School	NO	NO	3,34
	P2-5	Male	21	Anatolian High School	YES	NO	3,24
3	P3-1	Female	21	Anatolian High School	NO	YES	3,10
	P3-2	Male	23	Anatolian High School	NO	YES	3,86
	P3-3	Female	22	Anatolian High School	NO	YES	3,59
	P3-4	Female	21	Anatolian High School	NO	YES	3,61
	P3-5	Male	22	Anatolian High School	NO	YES	3,51
	P3-6	Female	22	Anatolian High School	NO	YES	3,16
4	P4-1	Male	22	Anatolian High School	NO	NO	3,43
	P4-2	Female	33	A High School in Turkmenistan	NO	YES	2,97
	P4-3	Female	21	Basic High School	NO	NO	3,36
	P4-4	Female	21	Anatolian High School	NO	NO	3,22
	P4-5	Female	22	Anatolian High School	NO	NO	2,63

Context

The present study were carried out with the teacher candidates studying at English Language Teaching Department of Gazi University. In this department, the students are trained according to the program determined by the Council of Higher Education (CoHE). However, the program was renewed by CoHE in 2018. Therefore, the freshmen students took the courses in the new program while the others continued to take the training according to the old one. Despite this change in the program, when the old and new programs are compared, it shows that both of them involve similar courses, as can be seen in Table 1.

However, there are also some courses that do not exist in the old or new program, and vice versa. For example, in the fall semester, the new program involves an Educational Sociology course given in Turkish while it does not exist in the old program. Similarly, in contrast to the old one, the new program has the courses of Critical Reading and Writing and World Englishes and Culture (elective), which are taught in the third semester. As to the fourth semester, two courses, English Language Teaching Curricula and Assessment of In-Class Learning (elective), are different courses from the ones in the old curriculum. Moreover, the Special Teaching Methods course is taught as two sequential courses in the fourth and fifth semesters of the old curriculum while this course does not exist in the new one. Finally, the Semantics and Language Teaching course is given as an elective in the eighth semester of the old curriculum while the new curriculum does not have it. Considering this comparison between the old and new curricula, it is possible to say that there are not big differences in terms of the courses. Therefore, it is expected that including the fourth grade students who took their training according to the old curriculum will not significantly affect the findings. Nevertheless, when compared with the situation where all participants are trained according to the new curriculum, this can be acknowledged as a limitation for this research.

Table 2

Comparison of the Old and New Curricula

Semester	Common Courses	
	Old Curriculum	New Curriculum
1 st Semester	Advanced Reading and Writing I, Listening and Pronunciation I, Introduction to Educational Sciences, Information Technologies in Education, Oral Communication Skills I,	Reading Skills I, Writing Skills I, Listening and Pronunciation I, Information Technologies, Oral Communication Skills I,
2 nd Semester	Advanced Reading and Writing II, Listening and Pronunciation II, Oral Communication Skills II, Educational Psychology	Reading Skills II, Writing Skills II, Listening and Pronunciation II, Oral Communication Skills II, Educational Psychology
3 rd Semester	English Literature I, Teaching Principles and Methods (Turkish), Approaches in Teaching English I, Linguistics I	English Literature I, Teaching Principles and Methods (Turkish), Learning and Teaching English Approaches, Linguistics I
4 th Semester	English Literature II, Linguistics II Language Acquisition, Scientific Research Methods, Approaches in English Language Teaching	English Literature II, Linguistics II, Language Acquisition, Research Methods in Education, New Approaches in English Language Teaching (Elective)
5 th Semester	Teaching English to Young Learners I, Teaching of Language Skills I, Literature and Language Teaching I, Classroom Management, Drama	Teaching English to Young Learners I, Teaching of English Language Skills I, Teaching of Language and Literature I, Drama in English Language Teaching
6 th Semester	Teaching English to Young Learners II, Teaching of Language Skills II, Literature and Language Teaching II, Measurement and Evaluation in Education (Turkish)	Teaching English to Young Learners II, Teaching of English Language Skills II, Teaching of Language and Literature II, Measurement and Evaluation in Education (Turkish)
7 th Semester	Teaching of Integrated Language Skills, School Experience, Special Education,	Teaching of Integrated Language Skills (Elective), Teaching Practice I, Special Education and Mainstreaming,
8 th Semester	Testing and Evaluation in English Language Teaching, Teaching Practice, Conversation Analysis and Language Teaching (Elective)	Exam Preparation in English Language Teaching, Teaching Practice II, Conversation Analysis and Language Teaching (Elective)

Data Collection Procedure

Before beginning to collect the data, the researcher applied for the ethical approval to Gazi University Ethics Committee and got the required permission on September 8th 2020. Then

the data collection process started, and it lasted throughout the 2020-2021 academic year. As Giraldo (2018) suggests for the LAL research, the data were collected through qualitative data collection methods. Therefore, the researcher used semi-structured and focus group interviews, opinion journals, scenarios, course observations, and documents as data collection methods.

At the beginning of the fall semester, the researcher analyzed the contents of the courses in the program by using the Information Package section on the website of Gazi University. According to this analysis, the researcher determined the courses that might involve content related to assessment in general or language assessment in particular. Then the researcher sent emails to the professors of the courses, collected the syllabuses, and observed some courses by participating in the lessons or watching the recordings on Perculus, which was an online platform supported by Gazi University for covering the courses in the pandemic. Moreover, piloting of the interview questions, opinion journals, and scenarios was done. In the spring semester, semi-structured and focus group interviews were held. The participants wrote their opinion journals and answered the questions asked in the Scenarios for Foreign Language Assessment (SfFLA). Moreover, the researcher observed the courses by participating in the lessons or watching them on Perculus and collected the documents that involved the syllabuses, lesson plans, and assignments. Moreover, the participants wrote opinion journals to answer the questions sent by the researcher. It should also be noted that the researcher sent a gift card from an online bookstore to the teacher candidates participating in the piloting or actual study after collecting the data.

Table 3

LAL Components and the Data Collection Tools

LAL Components	Data Collection Tools
KNOWLEDGE	Opinion Journal II Semi-structured Interview I and III Focus group Interview I
PRINCIPLES	Opinion Journal III Semi-structured Interview II Focus Group Interview II Scenarios for Foreign Language Assessment
SKILLS	Opinion Journal IV Semi-structured Interview III Focus Group Interview III Lesson Observations Documents (Lesson plans and Assignments) Scenarios for Foreign Language Assessment

Table 4

Data Collection Procedure

SEMESTER	DATE	PROCEDURE
FALL SEMESTER	October 5 2020- January 15 2021	Collecting course syllabuses and lesson plans Course observations Piloting of interview questions, opinion journals, and SfFLA
SPRING SEMESTER	March 1-7 2021 March 8-14 2021 March 15-21 2021 March 22-28 2021 March 29 – April 4 2021 April 5-11 2021 April 12-18 2021 April 19-25 2021 April 26 – May 2 2021 May 3-9 2021 May 10-16 2021 May 17-23 2021 May 24-30 2021 May 31 – June 6 2021 June 7-13 2021 June 28- July 2 2021	Announcement of the Research 1 st Journal (Experiences) + 1 st Interviews (Knowledge) 1 st Interviews 1 st Interviews 2 nd Journal (Qualifications) 1 st Group Interviews (1 st & 2 nd Journals + 1 st Interview) 3 rd Journal (Ethics) 2 nd Interviews (Ethics) 2 nd Interviews 4 th Journal (Skills) 2 nd Group Interviews (3 rd Journal + 2 nd Interview) 3 rd Interviews (Skills) 3 rd Interviews 3 rd Interviews SfFLA 3 rd Group Interviews (4 th Journal + 3 rd Interview + SfFLA)

Semi-structured Interviews

In the spring semester of the 2020-2021 academic year, one-on-one semi-structured interviews were conducted with each pre-service teacher in three rounds. During each interview, the participants were asked questions regarding one of the three components of LAL: knowledge, skills, and principles. These questions were prepared by taking into account the literature on LAL (Fulcher, 2012; Giraldo, 2018). The questions asked in the first round required the participants to define some basic terms related to assessment, such as test, assessment, reliability, validity, etc. The aim of asking these questions was to detect how much the participants knew about assessment. These knowledge questions were asked to the teacher candidates both at the beginning and end of the data collection process because, at the beginning of the semester, the teacher candidates in the 3rd and 4th grades had not taken the courses of Measurement and Evaluation in Education and Measurement and Evaluation in English Language Teaching, respectively. Therefore, by asking the same questions again after a long interval, it was aimed to get more robust results about the participants'

knowledge about assessment. Furthermore, I should note that I did not say the participants that I would ask the questions again. Therefore, the participants could not get prepared before the interview.

Before beginning these interviews, the researcher stated that it was expected that the participants to answer the questions by drawing on their knowledge if they have. However, they were also allowed to make comments if they did not have any specific knowledge about the terms or subjects mentioned in the questions.

As for the second round of interviews, it involved the questions addressing the *principles* component of LAL. That is to say, the participants answered the questions that were asked to elicit the answers regarding their awareness of an ethical assessment process and how to assure ethics in assessing their prospective students' language skills and components. In the third round of the interviews, the questions were about which methods, techniques, or tools the participants would employ to assess their students' language skills and components. Moreover, how they would score the assessment results and interpret and use them were also asked the teacher trainees through various questions.

Before each interview, the researcher read the journal written by each participant and prepared some questions unique to the participants about the issues, if any, that are vague or need to be detailed. Then, the researcher asked these questions to the participants at the beginning of the interviews for collecting in-depth data. After this process, the participants answered the questions prepared for the semi-structured interviews. During the interviews, the participants preferred to answer the questions in Turkish during all the interviews except just one interview although they were reminded that they could speak in Turkish or English. All of these interviews were conducted on ZOOM because of the Covid 19 pandemic and all of them were audio and video recorded by getting the permission of the participants. While conducting these interviews, each question was shared on the screen as a slide show with the help of screen sharing feature of ZOOM. The aim for this was to facilitate the participants to understand each question by seeing and hearing it.

Table 5

Duration of the Semi-structured Interviews

Grade	Participant	1 st Interview	2 nd Interview	3 rd Interview	TOTAL
1	P.1.1	73'	82'	100'	255'
	P.1.2	67'	56'	66'	189'
	P.1.3	53'	46'	86'	185'
	P.1.4	66'	91'	109'	266'
2	P.2.1	115'	117'	112'	344'
	P.2.2	80'	56'	108'	244'
	P.2.3	85'	56'	79'	220'
	P.2.4	52'	64'	77'	193'
	P.2.5	55'	91'	95'	241'
3	P.3.1	57'	70'	83'	210'
	P.3.2	70'	70'	114'	254'
	P.3.3	44'	43'	74'	161'
	P.3.4	63'	58'	69'	190'
	P.3.5	85'	70'	86'	241'
	P.3.6	51'	50'	93'	194'
4	P.4.1	70'	65'	132'	267'
	P.4.2	166'	105'	125'	396'
	P.4.3	47'	90'	91'	228'
	P.4.4	53'	69'	72'	194'
	P.4.5	63'	85'	96'	244'
TOTAL		1415'	1434'	1867'	4716'

Focus Group Interviews

In multiple-case studies, focus group interviews can be used for supporting the data from one-on-one interviews (Duff, 2008, p. 135). This type of interviews involve a small group of interviewees and are carried out with the aim of collecting a significant amount of data in an economical way (Dörnyei, 2007, p.144). Focus group interviews allow the participants not only to respond the questions asked by the researcher but also to express own ideas and to reflect on other responses (Ary, Jacobs, Sorensen, & Walker, 2014, p. 408). As a result, the interviewees can come to a consensus or disagreements may also occur because “consensus is neither necessary nor desired” (Fraenkel et al., 2023, p. 412). Focus group interviews typically involve four to six people (Cresswell, 2012, p. 218) and the interviewer asks five or six questions to the interviewees (Fraenkel et al., 2023, p. 412). As a disadvantage of these interviews, it can be referred that data transcription can be difficult because the transcriber may be unfamiliar to the interviewees’ voices and they may overlap (Duff, 2008, p. 135).

In this study, the researcher also conducted semi-structured focus group interviews in order to collect deeper data from all cases. In order to prevent data contamination, separate

interviews were conducted for each case. During these interviews, the researcher acted as a moderator who asked the questions to all interviewees one by one, encouraged the more silent interviewees to express their opinions, and summarized the agreed or disagreed points. Under these interview conditions, the participants generally came to a common decision while there were also some disagreements. Moreover, it is to note that all of the interviews were conducted in Turkish on Zoom and they were audio and video recorded. As it was in the semi-structured interviews, each question was shared on the screen as a slide show. Although some participants did not participate in some interviews, I asked the same questions to those participants in the later individual interviews in order to saturate the data. The questions asked in these interviews were prepared after the rough analysis of the data collected through semi-structured interviews and opinion journals. As Kelly (2003, p. 52) stated, focus group interviews can be used in conjunction with other data collection methods. The aim of it was to collect data to ensure data saturation. Furthermore, some questions which were prepared independently from the previous data and referring to some other issues were asked to the participants in these interviews.

Table 6

Duration of the Focus Group Interviews

Grade	1 st Focused Group Interview	2 nd Focused Group Interview	3 rd Focus Group Interview	TOTAL
1	52	94	54	200
2	64	97	65	226
3	55	93	81	229
4	51	78	68	197
TOTAL	222	362	268	852

Opinion Journals

The opinion journals consisted of four different journals, each of which was written as answers to two related questions asked for each journal. That is to say, the journal data were collected through eight open-ended questions in total. Of these questions, the first was about the participants' assessment experiences until the beginning of the data collection process. In conjunction with this question, the second one asked the participants what assessment practices they would change or do the same when they put themselves in their teachers' places. The aim of asking these two questions was to collect data about both the participants'

language assessment experiences, which have the potential to affect the development of their LAL (Berry et al., 2019; Yan et al., 2018), and their reflections on these experiences.

The questions in the second opinion journal were asked by targeting the *knowledge* component of LAL. These questions focused on the qualities that a teacher should have for conducting quality assessments and the potential factors that may influence the quality of assessments. These questions were asked because it is acknowledged that a teacher who is language assessment literate should be aware of the necessary knowledge about applied linguistics and language assessment (Scarino, 2013). In addition, s/he should be aware of such factors as context, assessment policies, stakeholders, and teacher beliefs that may affect the quality of assessments (Giraldo, 2018).

As for the questions answered in the third journal, they focused on the *principles* component of LAL. One of these questions asked the participants which ethical issues they would pay attention to while assessing their prospective students' language skills. The aim of asking this question was to learn whether the participants have concerns about such issues as fairness, transparency, bias, and democracy (Davies, 2008; Giraldo, 2018; Giraldo & Murcia, 2018). The other question was asking how they would ensure fairness among their students while conducting these assessments. Since it is acknowledged that a language assessment-literate teacher should "provide assessment practices that are fair and non-discriminatory", I asked this question (Giraldo, 2018, p. 190). It is also important to note that the participants were asked to write their journals in which language (English or Turkish) they felt they could express their opinions more easily and clearly. Eighty-four pages and 30,119 words of journal data were collected.

Lesson Observations

In order to collect data related to the assessment practices of the participants, the researcher participated in the courses or watched their recordings for two semesters and took notes about the points related to assessment and the practices made by the participants. Since each course did not involve microteaching, I determined beforehand which courses I would possibly observe. Then, I contacted the instructors of the courses and got permission to participate in their lessons or watch their recordings. These were the courses of Teaching English to Young Learners (TEYL) I and II, Teaching of English Language Skills (TELS) I

and II, and Teaching of Integrated Language Skills (TILS). While conducting these observations, I acted as an observer, not as a participant (Creswell, 1994, p. 149). Although Teaching Practice (TP) I and II courses also involved practices in the context of macro teachings, I did not observe them because I did not get permission from MoNE. However, I collected the lesson plans prepared by the participants in these courses.

Scenarios for Foreign Language Assessment

SfFLA aims to collect data about how the participants will assess their students when they are faced with the given situations. To this end, I prepared SfFLA, which presented different conditions that should be taken into consideration while answering the questions. While a great majority of these scenarios address the conditions for the assessment of the language skills of the students with different English levels, the remaining scenarios are related to diagnostic assessment, fair assessment, the effects of different assessment techniques on learning and teaching, and the washback effects of a high-stakes exam. While preparing the scenarios related to the assessment of language skills at different levels, the objectives in the MoNE's curricula for the 2-8 and 9-12 grades were used. The scenarios were initially prepared in English and then directly translated into Turkish. In order to ensure compatibility between the two forms, they were sent to an expert in the field of English language teaching, and the necessary revisions were made according to the feedback given by the expert.

Documents

The documents consisted of lesson plans, assignments, and syllabuses. The aim of collecting these documents was to ensure data triangulation. Because the first and second graders did not take any courses in which preparing a lesson plan was required, the lesson plans were collected from the third and fourth graders. These were the plans prepared for the courses of TEYL I and II, TELS I and II, TILS I and II, and TP I and II.

Data Analysis

The present research involves only the qualitative data which was collected through journals, interviews, scenarios, and document analysis. To analyze all these data, content analysis,

which is “a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use,” was conducted (Krippendorff, 2004, p. 18). The reason for doing a content analysis was “to obtain descriptive information about a topic” (Fraenkel et al., 2023, p. 434). While analyzing the data, the researcher adopted an inductive approach allowing the emergence of codes and categories from the data rather than being based on a predetermined scheme (Mayring, 2000).

In this qualitative content analysis, three main and eight sub-steps suggested by Selvi (2020), who adapted them from Elo and Kyngäs (2008), were followed. Accordingly, *preparation* is the first main step and includes the sub-steps: a) getting familiar with the data and research questions, and b) selecting the unit of analysis and making sense of the data. In this vein, the researcher became familiar with the data during the transcription process of the interviews. To this end, all interview data was transcribed verbatim by the researcher either by hand or using a computer program. That is to say, during the transcription process, the researcher transcribed a significant part of the data by listening to audio and writing verbatim what he heard. However, because this took too much time and the researcher had limited time in the doctoral program, the “Transcribe” feature of Microsoft 365 was used to transcribe the remaining data. It should be noted here that although this feature worked well for the most part, the researcher still listened to each audio recording and made the necessary changes to get the correct transcription. Furthermore, the researcher reviewed the research questions he had previously prepared and finalized them in the light of the relevant literature. As for the second sub-step, the researcher transferred all data to the MAXQDA 2020 Analytics Pro. This is a well-known qualitative data analysis program, and the researcher preferred to use it for two main reasons. The first and most important reason was that there was a large amount of qualitative data to be analyzed. Regarding this, Creswell (2012, p. 241) recommends the coders use computer programs for analyzing the data, which consists of more than 500 pages. The second reason was that the program had a user-friendly interface that involved four windows enabling the user to import data files, read each piece of data and assign codes to the relevant segments, and organize the codes under broader categories (Creswell, 2012, p. 243).

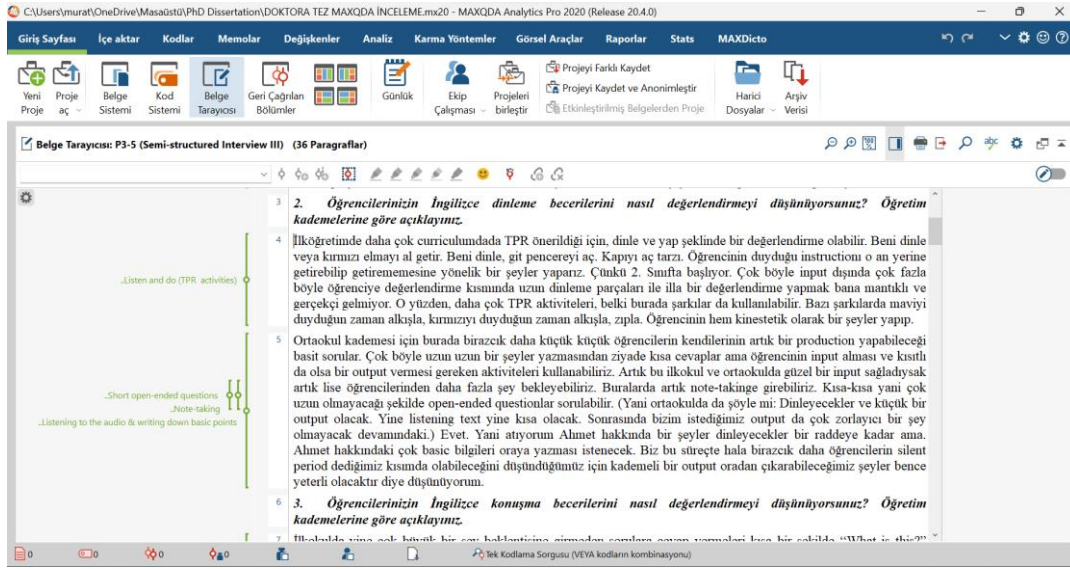


Figure 3. A screenshot of data analysis through MAXQDA Analytics Pro 2020

Organization is the second main step, and it can be conducted according to deductive or inductive approaches. In this research, the inductive approach, which allows that “the concepts are derived from data,” was adopted (Elo & Kyngäs, 2008, p. 107). With regard to this, there are five sub-steps for categorizing the data. Open coding is the first of them and it “means that notes and headings are written in the text while reading it” (Elo & Kyngäs, 2008, p. 109). For this, the researcher deployed the MAXQDA Analytics Pro 2020 and took some notes and assigned codes as in the Figure 3. This is followed by the coding sheets step and it involves putting the codes together. Since the MAXQDA is useful to gather similar or relevant codes under broader terms or categories, the categories were initially created and then the similar ones were collapsed under broader headings by comparing and contrasting with each other in order to ensure the step of grouping. In order to ensure inter-coder reliability, another researcher, who was pursuing doctoral degree as a research assistant in the field of applied linguistics, also coded a randomly selected sample of the data, which involved the thirty percent of the whole data. Before analyzing the data, we talked about the aim of the study and the research questions. Then I gave her the data in written form and she worked independently while coding the data. After she analyzed all the data, we discussed the codes until we reached a complete agreement.

Afterwards, these groupings were revised and some decisions were made to put some codes into one of the categories. Abstraction is the last step of the organization and the categories are finalized. Therefore, the emerging codes and categories were examined again and they were put into the main categories of weak or strong literacy.

Reporting of the results is the last of these three main steps. It involves the presentation of a model, conceptual map, or categories obtained at the end of the main data analysis. In this research, the findings are reported under the headings of the three main components of LAL: knowledge, skills, and principles. Additionally, the findings related to each of these components are presented under some sub-headings, and the categories of strong and weak literacy are given with the codes they involved in tables.

Pilot Study

A pilot study was conducted with eight pre-service English teachers before collecting the actual data. Two students, a female and a male, from each grade level (1st-4th grade) were the participants. Here, the aim was to find out how the interview questions, opinion journals, and scenarios were understood and answered by the English teacher candidates at different grades, and then to make the necessary revisions. To this end, the questions for opinion journals and SfFLA were initially sent to the participants via e-mail and asked to give the answers in detail without getting any help from any resources. Then, semi-structured interviews were conducted with the participants on ZOOM. At the beginning of these interviews, the researcher asked about the participants' vague answers and also elicited their reflections about the questions. Then, one-on-one semi-structured interviews were conducted, and at the end of each interview, the participants' reflections about the questions were received. According to these reflections, the researcher revised the data collection tools and got the opinion of an expert in the field of English language teaching.

Trustworthiness

In order to enhance the trustworthiness of this research, the researcher utilized a checklist created by Elo et al. (2014) for the field of nursing science and also suggested by Selvi (2020, p. 449) for the studies of applied linguistics. This checklist, designed according to the three basic steps of qualitative content analysis and presented in Figure 4, consists of questions to be answered for each step.

Phase of the content analysis study	Questions to check
Preparation phase	Data collection method How do I collect the most suitable data for my content analysis? Is this method the best available to answer the target research question? Should I use either descriptive or semi-structured questions? Self-awareness: what are my skills as a researcher? How do I pre-test my data collection method? Sampling strategy What is the best sampling method for my study? Who are the best informants for my study? What criteria should be used to select the participants? Is my sample appropriate? Is my data well saturated? Selecting the unit of analysis What is the unit of analysis? Is the unit of analysis too narrow or too broad?
Organization phase	Categorization and abstraction How should the concepts or categories be created? Is there still too many concepts? Is there any overlap between categories? Interpretation What is the degree of interpretation in the analysis? How do I ensure that the data accurately represent the information that the participants provided? Representativeness How to I check the trustworthiness of the analysis process? How do I check the representativeness of the data as a whole?
Reporting phase	Reporting results Are the results reported systematically and logically? How are connections between the data and results reported? Is the content and structure of concepts presented in a clear and understandable way? Can the reader evaluate the transferability of the results (are the data, sampling method, and participants described in a detailed manner)? Are quotations used systematically? How well do the categories cover the data? Are there similarities within and differences between categories? Is scientific language used to convey the results? Reporting analysis process Is there a full description of the analysis process? Is the trustworthiness of the content analysis discussed based on some criteria?

Figure 4. Checklist for researchers attempting to improve the trustworthiness of a content analysis study. Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K. & Kyngäs, H., (2014). Qualitative content analysis: A focus on trustworthiness. *Sage Open*, 4(1), 1–10.

Preparation is the first phase of the qualitative content analysis. In this phase, the questions about the data collection method should be answered. Starting with the first question, I collected qualitative data through opinion journals, semi-structured interviews, scenarios, and documents. The aim here is to triangulate the data and to get as rich data as possible. As for the second question, I consider these as the best methods to collect rich qualitative data about the PSETs' LAL. Although there are some inventories used in the literature to measure teachers' LAL levels (Ölmezer-Öztürk, 2018, pp. 168-171) they do not provide data on pre-service teachers' theoretical knowledge about assessment and how to ensure ethics. Therefore, deploying these data collection methods provides detailed data on these issues. As an answer to the third question, I used the semi-structured questions in the interviews because I aimed to collect data regarding some specific issues like the participants'

knowledge about the key concepts in assessment or how they would behave in order to assess language skills or to ensure ethics in assessment. The fourth question is about the researcher's self-awareness. In this regard, Elo et al. (2014) suggest that interview questions should be tested beforehand and the researcher should reflect on how s/he behaves during the interviews. In line with these, the questions for both the semi-structured interviews and the other data collection tools were piloted with other PSETs as explained above. Also, regarding the researcher's self-awareness, I should mention that in the pilot study, I asked all the interview questions in a word file on one page. However, after conducting one or two interviews, I realized that preparing a slide for each question would be more useful for the participant to pay attention. Therefore, in other interviews, I asked the questions in a slide show. Pre-testing of the data collection method is the last item of this sub-step, and piloting of the data collection tools meets this criterion.

Sampling strategy is the second sub-step of the preparation phase, and it involves questions related to how the sampling of the study is determined. Regarding this, I should note that I selected the participants conveniently. Although purposive sampling is very frequently used in the content analysis research (Kyngäs, Elo, Pölkki, Kääriäinen, & Kanste, 2011), I had difficulty finding participants due to the Covid-19 pandemic. Therefore, I recruited participants by contacting students' advisors and lecturers, and by sending messages about the study to students I knew myself. As a result, I was able to recruit a total of 20 participants within two weeks of the study being first announced. However, as a problem arising from the appropriate sampling method, it is not possible to say that these participants are the students who will provide the best information for this study.

Selection of the unit of analysis is the last sub-step of the first phase. Sentences constitute the unit of analysis of this research. Obtaining information about the literacy of teacher candidates requires analyzing the sentences they make. Therefore, I analyzed at the sentence level rather than words or phrases.

Organization is the second phase including the sub-steps of categorization and abstraction, interpretation, and representativeness. The categorization and abstraction refers to how the categories were created as a result of the qualitative context analysis. In this study, as a result of data analysis based on the inductive approach (Elo & Kyngäs, 2008), two categories emerged for each LAL component. These are *weak literacy* and *strong literacy*. Among

these, the "*weak literacy*" category for the "*knowledge*" component of LAL means that the teacher candidates either have no knowledge or have misconceptualizations about the concepts specific to the field of assessment. On the contrary, the category of "*strong literacy*" means that the prospective English teachers have enough knowledge about relevant concepts to be able to define them or present their features. As for the "*skills*" component of LAL, the "*weak literacy*" category means that the PSETs aim to use assessment tools that are non-communicative, mechanical and inauthentic. In contrast, "*strong literacy*" means that the teacher candidates reporting that they will use assessment methods that are communicative and authentic. Finally, in terms of the "*principles*" component of LAL, "*weak literacy*" includes not being aware of the characteristics of an ethical assessment process and not knowing how to ensure ethics in the evaluation process or engaging in practices that have the risk of threatening ethics. In the context of the "*principles*" component of LAL, "*strong literacy*" means that the teacher trainees are aware of the characteristics of an ethical assessment process and address the practices that enable to ensure an ethical assessment.

Interpretation refers to how much the researcher interprets the data while analyzing it. Although some amount of interpretation is involved in any qualitative content analysis (Lindgren, Lundman, & Graneheim, 2020), the researcher should not misinterpret this information. Moreover, the data should accurately reflect the information provided by the participants. Therefore, taking these issues into consideration, the data presented by the participants were analyzed without taking them out of context as much as possible. In other words, what the participants said or wrote was not only the focal point in the data analysis, but was evaluated together with the information they provided in all interviews and other data collection tools. In addition to this, in order to prevent over interpretation and ensure inter-coder reliability, another coder who works as a research assistant and pursuing a doctorate in applied linguistics also coded a certain part of the data.

Representativeness is the third sub-step of the organization phase. It involves the questions about how trustworthiness and representativeness of the data as a whole is checked. In this context, the checklist developed by Elo et al. (2014) was used to check trustworthiness during the data analysis phase of this study. Moreover, in order to ensure the representativeness of the data, there was constant back and forth between all data collection tools during the data analysis.

Reporting is the third phase and it involves the sub-steps regarding the reporting of results and analysis process. In response to the questions in this section, the findings section of this study was written in detail and systematically for each case and quotations were included to ensure that the findings reflect the data collected. In addition, the sampling method, participants and data collected are explained in detail to enable readers to transfer the findings. In addition, the findings of this study are reported in a scientific language. The questions in the sub-step of reporting analysis process ask whether the data analysis process is described in detail and the trustworthiness of the data analysis is based on a set of criteria. As answers to these questions, it is important to note that the data analysis section of the study includes detailed explanations on how the data analysis was conducted. Moreover, the criteria in the checklist created by Elo et al. (2014) were considered in order to ensure the trustworthiness of the data analysis.

CHAPTER IV

FINDINGS

Case 1: First and Second Graders

First and second graders constitute the first case of the research. This case involved nine junior pre-service English teachers. Of these, P1-1 was a 19-year-old male. His teachers from the 4th to the 8th grade focused on teaching grammar, and they used fill-in-the-blank, open-ended questions, and true-false statements for assessment. Although he did not face any challenges in these exams thanks to his sufficient knowledge of grammar, he stated that his friends, who were able to speak more fluently and write by using more various words than him, took low scores from these assessments. For this reason, he wrote that it would have been better if the exams involving writing a composition and oral exams had been done to assess their listening and speaking skills.

Similarly, in the 9th and 10th grades, he had to be assessed with these traditional tools of assessment, and he mentioned that “one of my two exams consisted entirely of multiple-choice questions”. According to P1-1, these paper-based exams are not beneficial for students “when English is just being introduced”. He expressed that this is because they force them to focus particularly on the grammatical structures and vocabulary that would appear in the exams instead of “dedicating themselves to improving their English”.

However, P1-1 experienced the assessment of four language skills in the 11th and 12th grades of upper secondary school. His performance in each skill was measured out of 25. For speaking assessment, he talked about a given topic and answered the questions asked by the teacher. During these assessments, his teacher paid attention to fluency, vocabulary choice,

and grammar. For the listening assessment, he answered the questions after listening to the texts read by the teachers because of the technical problems with the computers in the classrooms. However, he criticized this situation because his teachers could not read the texts fluently and correctly. Instead of this, he wished the technical problems had been solved and he could have listened to audio voiced by native speakers. As for the assessment of reading skills, he answered the wh- and multiple-choice questions after reading the given text. Moreover, his writing skills were also assessed in the exams. For this, he was asked to write his opinions about a given topic in at least 100 words. Regarding these assessments, he pointed out that “[he] did not think there was a problem with the assessments of reading and writing skills”. However, he mentioned that his teachers had not taught listening and speaking, and “this was totally dependent on the students’ own skills”. He states that this should not have been like that in a school where he entered with 470 out of 500 points.

P1-2, the other male member of this case, did not generally experience various practices for assessing four language skills. Reading was the skill that was the focus of these practices while the assessment of other skills was generally ignored. These assessments of reading usually involved open-ended and multiple-choice questions about the main idea of the text and the characters and places in it. Also, he noted that the questions for making inferences from the text were asked. However, he criticized these questions because the reading assessment deviated from its purpose and shifted to assessing the student’s ability to interpret. Therefore, he would not have included the inference questions in the reading assessments.

Regarding the assessment on listening skills, he did not face any assessment in the primary and lower secondary schools. However, he noted that, though it was rare, his listening began to be assessed in the upper secondary school with the use of smart boards in the classrooms. In these assessments, he was required to either fill in the given blanks while listening to the audio or answer the multiple-choice questions after listening. He wrote that listening to the audio voiced by both native and non-native speakers of English would have been better for the lessons and assessments “because there are not only British and American accents in the world and other accents must also be understood”.

Rarely than listening, his writing skills were also assessed a few times in the upper secondary school. These assessments on writing, which took place at the end of the exam paper, consisted of writing a short text to introduce himself or describe an experience with him or

a place he visited. Furthermore, as a writing assignment, he wrote a couple of times about a specific topic. However, he pointed out that these assessments were not adequate because the topics were not suitable for the level. Therefore, he wished he had written essays, which would have been more challenging for students like him who were language majors. In addition, it is important to note that he received feedback on his writings, and the teacher also scored them.

Regarding the ignorance of speaking skills, however, he said: "I have never had any training in speaking. I have never had it in primary, lower secondary, or upper secondary schools. I got it for the first time at university". For the assessment of speaking skills, he thinks that they might have been asked to talk about a particular topic for a specified period of time without any preparation.

P1-3, who is a Turkmen participant, was assessed with multiple-choice questions and fill-in-the-blank statements. In order to improve her English language skills, she went to a private language course, and she experienced the assessments with 'open-ended' questions there. They were the written questions asking about what and when the characters in the stories did. However, she noted that she had not received training in listening, speaking, and writing skills. Therefore, she did not experience any assessment of these skills. When she considered her experiences, she accentuated that she would have used multiple-choice and open-ended questions, fill-in-the-blank statements, and observation if she had been her teacher.

P1-4, who is a 20-year old female teacher trainee, started learning English in the fourth grade. In the Opinion Journal I, she expressed that she had been exposed to the traditional tools of assessment for many years: "Since the fourth grade, like almost all students in the country, I have always been exposed to multiple-choice, open-ended, or fill-in-the-blank tests that measure knowledge of grammar and vocabulary". However, she noted that "since [they] have generally been assessed with multiple-choice or open-ended questions, a system based on rote memorisation exists subconsciously". Therefore, for example, while she did not know how to use structures such as "will" and "going to" in speech, she just realised that she could distinguish in which situations these structures were the correct choice when answering the assessment items on paper. Considering this situation, she said that if she had been one of his previous teachers, she would have used group assignments in which students made postcards or posters, and wrote and told short stories. Moreover, related to the teaching and learning of grammar structures, she stated that she would make use of songs for more

permanent learning rather than memorising grammar structures in a context-free way, remembering that she sang songs with her teachers in the preparatory class in upper secondary school. However, she also added that "inevitably I would have been inclined to behave the way I had experienced it" and thus she reported that the traditional assessment items would have also been in the assessments she would have conducted.

Their language assessment experiences were generally similar. Regarding this, they stated that their language skills were not assessed through the use of alternative or communicative language assessment methods. Instead, their English teachers preferred to employ traditional and paper-based assessment methods.

P2-1 also had similar assessment experiences. In primary and lower secondary school, the assessment focused on the written input given by the teacher, which he had to memorize. These assessments consisted of writing down the Turkish meanings of the given English words, filling in the blanks by listening to a dialogue, and answering the multiple-choice questions. For the upper secondary school, he wrote: "When I started studying English as a major, my different English language skills began to be tested". As the testing methods, fill-in-the-blank and multiple-choice questions for vocabulary and grammar; writing an original story or biography by using some specific words and structures; translating the text from Turkish to English or English to Turkish; open-ended and multiple-choice questions for reading comprehension; group talk about a randomly chosen topic; fill-in-the-blank, in-class discussion, and question and answer for listening were used. However, he emphasizes that "for all language skills, multiple-choice, open-ended, and fill-in-the-blanks were predominantly used as assessment methods".

P2-1 described these assessment methods as monotonous and desired to be overcome quickly and wrote that he would have used methods in which his students would have participated continuously and willingly. According to him, the reason for this is that "students should learn English not to pass exams but to be a user of it in the world". Therefore, "[he] would have tried to avoid multiple-choice questions, except for grammar assessment, as far as time limits allowed, and [he] would have made sure that each skill was assessed productively in its own context".

P2-2 experienced fill-in-the-blank items requiring him to choose the correct word for the blank among the ones given above. Moreover, multiple-choice questions and writing

grammatical sentences by using the given words were the items used for assessing grammar. Regarding these items, he wrote: "I believe these were not efficient at all due to the fact that most of them were easily distinguishable and heavily depended on memorizing the rules". Because he studied in a social sciences upper secondary school, he had one year of education in a preparatory class. In this class, he took "many types of exams, focusing on each skill on a relatively basic level". The exams involved writing paragraphs, answering the multiple-choice questions after listening to dialogue twice, and giving individual presentations. However, he was exposed to more non-productive assessment methods like cloze tests, odd one-outs, "picture definitions (e.g., showing a picture of a doctor and asking the job or showing a snowy picture and asking the season)", true-false statements about a paragraph in the exam, and fill-in-the-blank items from 9th to 12th grades.

Criticizing these non-productive and memorization-oriented assessments, P2-2 wrote that he would have added "creativity to the exams by allowing the students to create meaningful sentences themselves". To do this, he would have provided the learners with some specific words or structures and asked them open-ended questions "so that they could create meaningful answers without being restricted to a specific amount of given words". Additionally, he reported that he had not taken an exam in which he was asked to draw and write a short text according to the drawing. Therefore, he would have "asked [the young learners] to draw a park with people in it and describe it themselves". For upper secondary school students, he would have given them various scenarios to use if clauses and asked them to role play in groups according to the given scenario and the type of if clause. Moreover, he would have given the learners some pictures (e.g. pictures of seasons and moods) and asked them to draw a line between two related pictures and write sentences to explain how they were related to each other. He noted that, in addition to the revisions, he would also have had his students make presentations to assess their speaking skills.

P2-3, who was a female teacher trainee, wrote: "From my primary school years to my university years, the assessment of my English skills was always carried out in a similar way." In this vein, she experienced multiple-choice questions, fill-in-the-blanks, open-ended questions, matching items, written exams, performance assignments, and quizzes from primary school to university. In addition to these assessment methods, oral exams consisting of teachers asking the meanings of words were also administered at the upper secondary school. Regarding these assessment methods, she noted that she did not generally approve

of them and gave the reason that the multiple-choice and open-ended questions would lead to rote memorization and very temporary learning. Instead of these tools for assessment, she would have conducted assessments focusing on four language skills and thus the learners' would have had more permanent knowledge.

As for her experiences in upper secondary school, she wrote "When I think of an English teacher, I always think of my high school teacher, so I will answer this question based on her". She noted that she would use the same assessment methods as her teacher, stating that the teacher had to make written assessments according to the regulations. P2-3 emphasized that her teacher asked questions that students could express their own thoughts as well as knowledge questions, and she would ask not only grammar questions but also open-ended questions. Instead, she would have given her students tasks in which they could have used English as a conversational tool. For example, s/he would have organised discussion groups of three and asked them to discuss or make presentations on a topic given by her or chosen by the students.

Multiple-choice questions and fill-in-the-blank items were the assessment tools that P2-4 also had to answer in the written exams until she started university. Moreover, she added that she had not received enough feedback on her errors from her teacher. Taking these conditions into consideration, like P2-3, she also thinks that they cause the learners to memorize only vocabulary and grammar structures. Unlike the primary and lower secondary schools, however, she participated in the weekly speaking activities and made a presentation about a topic she chose in the upper secondary school.

P2-4 thought that the assessment methods he experienced until tertiary level were "inadequate, distracting the student from attaining objectives and from the educational environment," and directly rendered language learning passive. In this vein, she reported that if she were her English teacher, she would have organized activities in line with her students' interests and needs and in accordance with their age range. Moreover, she would have made her assessments in a way that would make her students more interested in the lesson. To exemplify, she would have initially detected the topics about which her students would want to talk and then asked them to express their opinions in small groups. While her students were speaking in English, she would have observed them and taken notes with regard to their performances. However, she would have carried out this as a practice of informal

assessment, and therefore no scores would have been assigned. Rather, she would have provided feedback to encourage her students to speak the target language.

P2-5 also took multiple-choice and fill-in-the-blank items in the written exams administered in the primary and lower secondary schools. However, like P2-2, he studied in a preparatory class in a social sciences upper secondary school and received assessments on four language skills. After this one year of education, he was subjected to paper-based assessments involving the classical testing items. As other PSETs did, he also emphasized their ineffectiveness on language learning with these words: "These exams, which did not contribute to my language skills, tried to make me memorize things rather than teach me a language."

P2-5, who described the non-communicative assessment methods he experienced as "wrong," wrote that the reason for using these methods was "teacher-centered lecturing." However, he stated that there had been nothing much that teachers could have done in this regard due to the limited facilities offered to them and the crowded classrooms. Nevertheless, he stated that his teachers could have tried to make assessments focusing on the four language skills. In this respect, he suggested that they could have used pen-and-paper exams for reading and writing (e.g., writing a story) and debates for listening and speaking.

Literacy in the Knowledge Component of LAL

Knowledge about the Key Concepts of Assessment

The findings presented in this section were related to the definitions of the five key terms regarding the field of assessment: test, assessment, measurement, and evaluation. Two categories emerged as a result of the data analysis from the first and third interviews: weak literacy and strong literacy.

Table 7

Knowledge about the Key Concepts of Assessment

Categories	Codes (Semi-structured Interview I)	Codes (Semi-structured Interview III)
WEAK LITERACY	<i>Test</i>	<i>Test</i>
	- Multiple-choice questions (P1-4, P2-4, P2-5) - Testing what someone knows through a test (P1-3) - A measurement system measuring whether the teaching materials are used appropriately for their intended purposes (P2-2)	- Multiple-choice questions (P1-1, P1-3, P1-4, P2-1, P2-4, P2-5) - Testing (P1-2) - Fill-in-the-blanks (P2-1) - Measuring (P2-4)
	<i>Assessment</i>	<i>Assessment</i>
	- Unable to define (P1-1, P2-4) - The aim of the test (P1-2) - Not knowing the difference between testing and assessment (P1-3) - Giving results about various aspects (e.g., content, style) (P1-4) - Evaluation, measurement, or level detection (P2-1) - Studies enabling the student to realize what s/he knows outside of the scoring and official exam format (P2-2) - Measurement based on determining the state of the information given to the student (P2-3) - Making interpretations about the data from testing and taking actions (P2-5)	- Scoring (P2-3, P2-4) - The method to measure the level of learning (P2-1) - Analysis of the level of student's language skills according to the results of testing (P2-5)
	<i>Measurement</i>	<i>Measurement</i>
	- Unable to define (P1-3, P2-1) - Similar to assessment (P1-2, P2-3) - Measuring something physically (e.g., mass) (P1-2) - Absolute numerical value (P1-4) - Involving testing and the student's emotions and psychology (P1-1) - The concept that includes both examination and non-examination tasks (P2-2) - Including the terms: Test and assessment (P2-2) - Measuring with a written document, a test, or a quiz (P2-3) - Assessment in a numerical sense (P2-5) - An achievement test on paper, in exam format (P2-5)	- Unable to define (P1-3, P2-1) - All assessment tools (P1-1) - Measuring a length and kilogram (P1-2) - A term involving "test" and "measurement" (P2-2) - Defining quality, knowledge, or objective by measuring according to some criteria (P2-3) - Additional methods to assessment (P2-4)
		<i>Evaluation</i>
		- Scoring the performance in assessment according to the criteria (P1-1) - Qualitative evaluation (P1-2) - Unable to define (P1-3) - Same with assessment (P1-4) - Umbrella term (Embracing all terms above) (P2-4) - The results of actions taken according to the results of the assessment (P2-5)
		<i>Reliability</i>
		- Unable to define (P1-2, P2-3) - Impartiality (P1-1, P2-4) - Not cheating (P1-1) - Trusting the students (P1-3) - Measuring everything (not) done by the student throughout assessment (P1-4) - An assessment process prepared according to some specific criteria (P2-1)

<i>Evaluation</i>	• The same with assessment (P1-3, P1-4, P2-3) • Scoring and interpreting outputs according to criteria (P1-1) • Determining the value of an object (P1-2) • Evaluating the use of different language components through assignments (P1-4) • Unable to define (P2-1) • A system scoring for how much theoretical knowledge is put into practice (P2-2) • The same with measurement (P2-3) • Score (P2-4) • Identifying the effects of the actions taken according to the interpretations of the data from the assessment (P2-5)	• The situation of the scoring criteria being unchangeable, precise, objective, and impartial (P2-2) • Not asking the student about what was not taught (P2-4) • How mathematically correct or incorrect the tests are (P2-5)
<i>Reliability</i>	• Impartiality (P1-1, P1-2, P2-2) • Teachers and students obeying the pre-determined rules (P1-1) • Unable to define (P1-3) • Teacher's consideration of the student's performances throughout a semester (P1-4) • An assessment method approved by some authorized persons or institutions (P2-1) • Assessment of what is taught (P2-3) • Teacher's adherence to the assessment criteria (P2-4) • The extent to which the teacher tests what s/he teaches to the students (P2-5) • Testing what should be tested (P2-5)	<i>Intra-Rater Reliability</i> • Unable to define (P1-2, P1-3, P2-1, P2-5) • Teacher consistently following the assessment criteria (P2-2, P2-4) • Fair assessment (P1-1) • Not having any doubt about the assessment done by the teacher (P1-4) • Impartiality of the person or institution doing the assessment (P2-3) • The consistency of the teacher (P2-4) <i>Inter-Rater Reliability</i> • Unable to define (P1-2, P1-3, P2-1) • Rater assessing according to a common principle (P1-1) • The trust among the English teachers (P1-4) • Assessing different performances according to the same criteria (P2-2) • The appropriateness of the student group to the aim and assessment tool (P2-3) • Not sharing the exam questions (P2-3) • Consistency between the teacher and the criteria (P2-4) • Changing according to the assessor (teacher or institution) (P2-5)
<i>Intra-Rater Reliability</i>	• Unable to define (P1-1, P1-3, P2-4, P2-5) • Fair assessment (P1-2, P2-2) • Assessment accomplishing its aim (P1-2) • Teacher's involving own feelings into scoring and evaluation (P1-4) • Measuring the exact and correct knowledge and focusing on the target topic to ensure reliability for the assessee (P2-1) • Assessor having the qualities to conduct the assessment (P2-3)	<i>Methods to Estimate the Reliability</i> • Not having any knowledge (P1-1, P1-2, P1-3, P1-4, P2-1, P2-2, P2-3, P2-4, P2-5) <i>Validity</i> • Assessment of the taught content (P1-2, P2-2, P2-3, P2-4) • Unable to define (P1-3, P1-4, P2-1) • Administration of an exam under the same conditions (P1-1) • The assessment being objective (P2-2) • The appropriateness of the assessment tool to the aim of the assessment (P2-3) • Assessment of the taught content in a recent period (P2-3) • Acceptance of the assessment by other people or institutions (P2-5)
<i>Inter-Rater Reliability</i>	• Unable to define (P1-3, P2-1, P2-2)	<i>Types of Validity</i> • Not having any knowledge (P1-1, P1-2, P1-3, P1-4, P2-1, P2-2, P2-3, P2-4, P2-5) <i>Practicality</i> • Clearness of the instructions in the exam (P1-1) • Related to the quality of the assessment method (P1-2) • Teacher using different types of questions for each assessment (P1-3) • The assessment allowing the students to perform their language skills (P2-2) • The appropriateness of the assessment tool to the aim of assessment (P2-3) • Tests giving accurate results (P2-5)

- Different teachers giving equal information to students about assessment (P1-1)
- Different assessment methods accomplishing the same goal (P1-2)
- Teacher and student being honest and sincere with each other (P1-4)
- Confidentiality of the information received from the students among the raters (P2-3)
- Teacher consistency across different criteria (P2-4)
- The difference among the assessments conducted by different teachers (P2-5)

Methods to Estimate the Reliability

- Not having enough knowledge (P1-1, P1-2, P1-3, P1-4, P2-1, P2-2, P2-3, P2-4, P2-5)

Validity

- Unable to define (P1-2, P1-3)
- Student's products being of acceptable quality (P1-1)
- Being un/successful according to some criteria (P1-4)
- No cheating (P2-1)
- No breaking the rules (P2-1)
- Approval of the assessment method by some authorities (P2-1)
- Suitability of the assessment group for the aim of the assessment (P2-1)
- An assessment method where certain criteria should be met (P2-1)
- Every student passing the set limit criterion (P2-2)
- Invalid assessment if the majority fails (P2-2)
- Consideration of the assessment according to the conditions of the era when it takes place (P2-3)
- Teacher giving the students criteria regarding the assessment (P2-4)
- Teacher adherence to the criteria and not considering other criteria than the given ones (P2-4)
- International acceptability of exams (TOEFL, IELTS versus FLT, FLPT) (P2-5)

Types of Validity

- Not having enough knowledge (P1-1, P1-2, P1-3, P1-4, P2-1, P2-2, P2-4, P2-5)

Washback

- Unable to define (P1-1, P2-1, P2-2, P2-3, P2-4)
- The effects of feedback on the students' learning (P1-4, P2-5)
- Assessment assessing itself (P1-2)
- Detecting how much the student learned according to the assessment results (P1-3)

	• Validity according to time (P2-3) • Validity according to the group of assessments (P2-3) • Validity regarding knowledge (P2-3) • Validity regarding the rater (P2-3) <i>Practicality</i> • Positive effects of assessment types on students' learning (P1-1) • Related to the quality of the assessment method (P1-2) • Unable to define (P1-3) • The measurement tool giving valid results for each student (P1-4) • Usefulness of the assessment results (P2-1) • The student using the knowledge s/he used in the assessment in real life (P2-2) • Appropriateness of the assessment type in terms of the purpose and the person/group to be assessed (P2-3) • Clarity of the instructions (P2-4) <i>Washback</i> • Unable to define (P1-3, P2-1, P2-2) • Student making errors in assessment due to incorrect teaching (P1-1) • The negative effects of the traditional education system on language learning (P1-1) • Methods aiming to evaluate the quality of teaching (P1-2) • Student reflecting the teacher's behaviors to the teacher (P1-4) • The effects of the assessment results (P2-3) • The assessment results giving information about the success of teaching (P2-3) • The positive or negative effects of the score the student got on the subsequent assessments (P2-4) • The actions the student should take according to the assessment results (P2-5) • The demotivating effect of the assessment results (P2-5)	
STRONG LITERACY	<i>Test</i> • Tools or procedures used to detect whether the student acquired a skill or knowledge (P1-1, P1-2, P2-1, P2-3) <i>Practicality</i>	<i>Test</i> • A tool requiring the students to perform their language skills, generally written (P2-2)

-
- Teacher assessing the students without being tied to place and time (P2-5)
 - Videotaping yourself (practical) (P2-5)
 - Appropriate to use without the risk of cheating (P2-5)
 - Written exams-Not practical-The need for proctoring (P2-5)
 - Debate: Practical-No need to proctor and less responsibility (P2-5)

- A document or tool involving the questions or criteria based on the quality or skill to be assessed (P2-3)
- Involving every type of question depending on the quality to be assessed (P2-3)

Assessment

- Collecting information regarding the proficiency in the language skills (P1-2, P1-3, P1-4)
- Administering exams and assignments to test how much the student learned and could use it (P1-1)
- Process-oriented practice involving the use of assignments and tasks (P2-2)

Measurement

- Score given at the end of the assessment (P1-4, P2-5)

Evaluation

- Making interpretations on the assessment results by comparing them with some criteria (P2-1, P2-2, P2-3)

Practicality

- Feasible (P2-1, P2-4)
 - Time-saving (P1-4, P2-4)
 - Easy to score (P1-4)
 - Economic (P2-1)
-

Weak Literacy of the Key Concepts of Assessment (First Semi-structured Interview)

The junior PSETs gave incorrect definitions regarding the key terms or could not provide any answers to define them. The terms like test, assessment, measurement, evaluation, reliability, validity, practicality, and washback were the ones about which they were not adequately knowledgeable. For example, some of them stated that the concept of “test” meant multiple-choice questions, as in the excerpts below:

I mean, a paper with 10 to 15 questions, each with 4 to 5 choices in the form of A, B, or C; in other words, a test, an evaluation, an exam, that kind of thing (P1-4, Semi-structured Interview I).

I cannot fully define the concept of “test”, but the first thing that comes to mind is a multiple-choice question (P2-4, Semi-structured Interview I).

As for the assessment, two of the participants (P1-1 and P2-4) could not provide any explanation or definition regarding it, while the others made various guesses or comments. According to P1-2, it refers to the aim of the test, while P1-4 defines it as a type of practice that is comprehensive and gives results about more than one feature. As an example of this, she commented that it can reveal information regarding the content, style, or presentation of a project. Unlike this, P2-1 remarked that assessment involves measurement, evaluation, and detecting the student's level in a subject. P2-2, on the other hand, defined the concept as an informal practice enabling students to realize what they know or do not know. P2-3 considered assessment as a result obtained to determine how much the learners learned from what the teacher taught them. P2-5's definition of assessment involves the interpretation of results from testing. This interpretation will allow him to make informed decisions and take the necessary actions for the success of subsequent instructions.

Measurement is a key term that none of the PSETs in Case 1 had accurate knowledge about. Regarding this, two pre-service teachers (P1-3 and P2-1) could not make any comments or guesses, while the others said what it evoked in them based on its name or the order in which the concept was asked. Moreover, it should be noted that four participants made two different comments related to measurement. In their answers, they confused measurement and assessment concepts or commented that it involves other concepts such as test and assessment as an umbrella term. Also, they drew on the word “measure” and guessed that it meant measuring just weight or height. Tests and tasks for assessment were also related to the term “measurement” by some PSETs.

Evaluation, like measurement, is another concept that no PSET could define correctly. While three of them stated that evaluation is the same as assessment or measurement, some others associated it with scores, scoring, or determining the value of an object or the effects of actions. To illustrate, P1-2 addressed the Turkish equivalents of the three concepts, assessment, measurement, and evaluation, and defined evaluation as determining the value of something, as in the following excerpt:

If I want to distinguish evaluation a little bit, I would say at least in the sense of determining the value of something. Apart from that, it doesn't really distinguish itself when translated into Turkish. [...] Actually, there is a difference, but it doesn't make much difference when translated into Turkish. They are all translated as evaluation. Evaluation, for example, can be used to determine the value of an item. Measurement, on the other hand, means measuring something physical, such as mass, kilograms, height, and so on. Assessment is related to the subject we are interested in. It is a bit more in education. I can distinguish in this way. But other than that, it is almost the same when we translate it into Turkish (Semi-structured Interview I).

In addition to these basic terms in assessment, the junior teacher trainees were not also knowledgeable about the two main principles of assessment: reliability and validity. Of these, with regard to reliability, all but one participant (P1-3) had a guess about what it was. However, while three of the participants (P1-1, P1-2, and P2-2) defined it as the teacher being impartial during the assessment process, the others referred to different points in their answers. To illustrate, P1-4 mentioned reliability as the teacher tracking the student's language development by observing all their performances throughout a semester, while P2-1 commented that it means the acceptance of an assessment method as qualified by specific institutions and experts.

Consistent with their limited literacy in reliability, the participants were also unable to accurately define the concepts of intra-rater reliability. Regarding this concept, four of them could not provide any definition, while two participants (P1-2 and P2-2) mentioned the fairness of the assessment. In contrast to this, P1-4 defined intra-rater reliability as an unfair assessment situation in which the teacher included his or her emotions in the assessment process. Unlike these, P2-1 explained this concept as the situation where the assessment measures exactly the targeted knowledge without going beyond it, and he added that the aim of it is to create trust in the assessee. P2-3, who analyzed the concept from another perspective, commented that it means the characteristics that a teacher should have in order to make qualified assessments. Moreover, it is important to note that the PSETs in Case 1, as expected, do not have any prior knowledge about reliability estimation methods.

As for the inter-rater reliability, none of the junior PSETs could define or explain it accurately. To be more precise, three of them could not make any comments or guesses about it. On the other hand, the other participants gave different explanations. For example, one of them defined it as teachers giving information to students about assessment, while another one emphasized the purpose of assessment and explained it as different assessment methods serving the same purpose. In addition, one participant defined it as the relationship between the teacher and the student, while another defined it as the ethical storage of personal information about students between raters and therefore not sharing it with others. According to P2-4, this concept means that a teacher acts consistently between the different criteria used in assessment. P2-5, on the other hand, considered this concept from a different perspective and defined it as the inconsistency of different assessments done by different teachers in terms of quality and content.

Validity, which is another very important principle of assessment, is another concept about which the participants did not have accurate knowledge. Regarding this, seven members of Case 1 could provide answers by mentioning various subjects that they thought were important for assessment. P2-1, for instance, made more than one guess about validity. His guesses were related to the security of the assessment, the acceptance of the assessment by some institutions or experts, the assessment method, including specific criteria to be met, or the appropriateness of the assessment group to the aim of the assessment. Thinking in the same logic as P2-1, P2-5 also considered validity in terms of national and international exams and commented that exams such as TOEFL and IELTS have validity while exams such as the Foreign Language Test (YDT) and the Foreign Language Proficiency Test (YDS) do not. As understood from this interpretation, according to P2-5, validity means the international acceptance of a high-stakes test. When the answers given by the other participants are considered, four of them refer to the existence of some criteria or the comparison of the students' products, performances, and scores with these. For example, P1-1 defined validity as the quality of the products prepared by students using language and stated that if these products are above a level determined by the teacher, they will be called valid. However, P2-3 considered this concept from a different point of view and interpreted it as the assessment being in accordance with the conditions of the age.

I guess I can interpret it like this: Everything has a period of validity. It should be analyzed according to its period. Information is constantly changing. We cannot say that the information is certain. An assessment made 100 years ago cannot be equated with today. It is also a related

concept. I define this as meaning that in order for the assessment to be valid, it needs to be evaluated again according to its period (Semi-structured Interview I).

P2-3 was the only participant to respond regarding the types of validity. Her guesses about these types were based on what the words "validity" and "assessment" evoked in her. In this vein, she said that validity is categorized into types according to the circumstances of the age, the characteristics of the students targeted by the assessment, the knowledge that the assessment aims to detect in students, and the qualities of the assessor.

Practicality and washback are other key terms that the junior PSETs are not familiar with. They were unable to characterize the first one of them clearly, including a number of features such as economy, feasibility, and time efficiency (Brown, 2004, p. 19). Instead, they discussed such issues as the effects and quality of the assessment, the compatibility of the assessment with real-life situations, the clarity of the instructions, and the usefulness and validity of the assessment results.

As for the washback, which means "the influence of testing on teaching and learning" (Cheng & Curtis, 2004, p. 4), none of the junior student teachers knew what it was. In other words, three of them said that they did not have any comments or guesses about the concept, while the others made some definitions that were not compatible with the actual meaning of washback. These definitions were about the effects or quality of teaching or the education system, the relationship between teacher and student, and the effects of the assessment results. To illustrate, P1-1, in his second definition, addressed the concept of washback in the macro dimension and explained how a traditional education system would affect the language learning process as follows:

In fact, this is something that can also be addressed in the education system. For example, an education system works completely with traditional methods. In the future, this student will become an individual and will speak English in his/her future life, for example (I addressed the questions from English teaching). It will result in discomfort. He will experience the effects of this. In my opinion, the washback effect continues like this if I address it more broadly (Semi-structured Interview I).

Strong Literacy of the Key Concepts of Assessment (First Semi-structured Interview)

Some junior student teachers were also able to provide some correct answers to define the two terms test and practicality. Regarding test, they gave the definition that they refer to the

tools, methods, or procedures exploited to assess the students' skills or knowledge related to a subject area. To exemplify, P1-1 explained what the test was in the following terms:

A test is a procedure to evaluate an individual. This can be an exam paper or a process that can be done for that individual. It is a set of procedures performed to get information about that individual or to evaluate the individual in every subject you can think of (Semi-structured Interview I).

P2-5 was the only participant who could give an accurate explanation about practicality. In his answers, he focused on the feasibility of assessment techniques or tools. Stating that a practical assessment technique can have a mobile feature rather than being tied to a place or a specific period of time. To exemplify this, he stated that the students videotaping themselves will be a practical practice for speaking assessments since they can do that at any place or time. Additionally, he added that this technique should also not be conducive to cheating. Therefore, he commented that written exams are not practical, while debates are practical assessment techniques because they do not require proctoring.

Weak Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)

The members of the first case defined the terms asked in the interviews incorrectly, or they stated that they could not define them. These incorrect definitions were given for such terms as test, assessment, measurement, evaluation, reliability, validity, practicality, and washback in the third semi-structured interviews. For example, while defining the term "test", some of the participants referred to it as an exam paper consisting of multiple-choice questions as in the following excerpts:

When I consider the concept of test in the context of education. When I think of test, something comes to mind. While evaluating the student, waiting for the student to choose the correct answer according to the question by giving 4, 5, 3, or more multiple choice questions and answers is the test here (P1-1, Semi-structured Interview III).

Re-requesting the information given to the student in a somewhat test-like format. It comes to mind as choices (P2-5, Semi-structured Interview III).

Moreover, P2-1 expressed that fill-in-the-blank items were also another component of the "test" in addition to the multiple-choice questions. As different from these definitions, P1-2 defined this term as the act of administering a test while P2-4 pointed out that it was "measuring" the learners' knowledge.

As for the definition of assessment, P2-1, P2-3, P2-4, and P2-5, gave incorrect definitions. For example, P2-3 defined the term “assessment” as assigning scores according to the students’ success or skill and accentuated that its aim was to obtain these scores and analyze them. Likewise, P2-4 mentioned that the assessment involved the criteria for measuring the learners’ skills and added that scoring according to some specific criteria was the main feature of the assessment. As different from these participants, P2-1 referred to assessment as the methods adopted in order to measure the students’ knowledge. On the other hand, P2-5 also made a different definition and stated that the assessment meant the level of the learners’ language skills and knowledge determined as the result of testing.

Measurement also could not be accurately defined by the junior PSETs. While two of them (P1-3 and P2-1) could not make any comment on this term, some others tried to define it but could not give an accurate definition. For example, P1-1 guessed that “it may be tests and oral exams”. P1-2 drew on the meaning of the word “measure” and stated the term was related to measuring something physically, like length and kilograms, rather than quality. P2-2 defined measurement as an umbrella term involving test, testing, and assessment. According to P2-3, measurement is defining a piece of knowledge, an attribute, or an achievement by measuring it in certain scopes. As different from this, P2-4 commented that measurement refers to the methods used as additions to the basic assessment, and these allow the teacher “to see whether the student has really learned or not”. In addition, she adds that “there is no direct grading or evaluation in this method as in the assessment”.

Evaluation is also another term that the junior PSETs could not define accurately. While P1-3 wanted to skip it by stating she did not have any idea about it, the other four teacher trainees made different comments about the term. For example, P1-1 related it to scoring and said, “This can be the scoring of the things we assess with tests, oral exams, or any form of speech”. According to P1-2, it is “a qualitative assessment” of the students’ success rather than a quantitative measurement, as in the explanation below.

This is a qualitative assessment. On the basis of quality. I think of them as evaluating a certain thing, such as quality, success, etc. Measurement is what we call measuring objective concepts that everyone accepts. For example, Ayşe's or Fatma's height or weight, we call these as measurement. But evaluation is Ayşe's success in volleyball. How good is she at volleyball? Or evaluations such as Fatma's success in mathematics (Semi-structured Interview III).

While P1-4 stated that there was no difference between assessment and evaluation, P2-5 pointed out that they referred to two different situations. According to P2-5, evaluation is

the judgment about the effectiveness of the actions taken based on the analysis made from the assessment results.

When it comes to the terms related to the principles of assessment, the first graders either made erroneous guesses about them or were unable to comment at all. Reliability, which is one of the desired attributes of assessment results, is a term about which the participants do not have any background knowledge. Regarding reliability, P1-1 commented that it is a term that involves ethical values such as the prevention of cheating and the teacher's impartiality in his or her assessments. While P1-2 and P2-3 had no answer, P1-3 defined it as the teacher's confidence in the students that they did not cheat during the exam. On the other hand, P1-4 pointed out that reliability might emerge thanks to the assessments, which were longitudinal and comprehensive, as explained in the excerpt below.

I think it's all about the assessment, which is clear about what's going on end-to-end, from A to Z. So those long, comprehensive assessments provide reliability. In other words, without them, nothing very ethical or very reliable can happen. So reliability is something that emerges thanks to it. [So is it something like measuring everything? Full, large scale.] It is not dynamic. It is not dynamic, but, for example, what did the student really do during that period? What she did or did not do, what she missed, what she made an effort for, and measuring all of them and scoring provide reliability (Semi-structured Interview III).

P2-1 commented that reliability meant that the assessment process was prepared according to specific criteria. In a similar vein, P2-2 mentioned the scoring criteria for defining reliability and stated that reliability was the situation of these criteria being unchangeable, precise, and objective so as not to cause any partiality among the students. P2-4 made two different interpretations of the concept of reliability from the perspectives of teachers and students. While she made the comment for the student that the assessment should not be about what was not taught, she expressed for the teacher that the teacher should be objective and consistent in awarding scores. As for P2-5, he defined reliability as how mathematically correct or incorrect the assessments made to the student actually were.

Intra-rater and inter-rater reliability, which are about the consistency of the scores given by the same rater or at least two raters, are the other concepts that the first graders are not familiar with. While defining intra-rater reliability, P1-1 referred to the fairness of the assessment. He defined it as the assessor's assessing the learners in accordance with the criteria in his or her mind without any partiality and not including his or her own feelings. That is to say, the concept of intra-rater reliability reminded P1-1 of an assessment appropriate to ethical principles. While P1-2, P1-3, P2-1, and P2-5 did not have any guesses

about the term, P1-4 commented that the intra-rater reliability sounds like the student's not having any doubt about the assessment done by the teacher. While P2-2 and P2-4 defined the intra-rater reliability as the situation of the teacher's not going beyond the pre-determined assessment criteria, P2-3 guessed that it was related to the teacher or institute as the sources of assessment being impartial to the students.

As in the intra-rater reliability, the participants could not give an accurate definition of the term "inter-rater reliability". For example, it did not mean anything to P1-2, P1-3, and P2-1, and therefore, they wanted to skip it. In contrast to this, P1-1 remarked that it was the teachers communicating with each other and conducting their assessments according to a common principle. While she was not very sure about it, P1-4 accentuated that it referred to the trust among the English teachers assessing their students. According to P2-2, the use of the same criteria for assessing the students' different performances means inter-rater reliability. P2-3 made two guesses about the term. While one of them was about the compatibility between the group of learners and the goal and tool of assessment, the other was about the confidentiality of the exam questions. P2-4 commented that the consistency between the teacher and the assessment criteria meant inter-rater reliability. P2-5 guessed that the term might refer to the difference between the sources of assessment and gave the example that it might change among the teachers or "exam centers like MoNE and Council of Higher Education". In addition, it should be noted that the junior PSETs did not have any knowledge about the strategies of estimating reliability, and therefore they could not make any comments about them.

Validity was another crucial term that the members of the first case failed to define. Of them, P1-3, P1-4, and P2-1 could not give any explanation about their validity. On the other hand, validity was most frequently defined by the teacher candidates (P1-2, P2-2, P2-3, and P2-4) as the assessment of the content covered in the course. As different from this definition, P1-1 also approached validity from an ethical point of view, as he did with reliability, and he gave an explanation by drawing on the high-stakes exams.

I'll discuss it for the high-stakes exams. We talked about these ethical principles in the big exams for both students and teachers. Now I'm thinking of a much larger version of it. In order for an exam to be valid, I think of principles such as applying an evaluation criterion to everyone in the same way under the same conditions, not allowing any outside intervention, and allowing everyone to take it at the same time with the same tools (Semi-structured Interview III).

In addition to this definition, P2-2 and P2-3 made other guesses about the term. That is to say, while P2-2 stated that validity was the situation of assessment being objective, P2-3 remarked that validity might be the appropriateness of the assessment tool to the aim of assessment and the assessment of the taught content in a recent period of time. As different from these definitions and explanations, P2-5 drew on the word “valid” and remarked that validity meant the acceptance of an exam by other people or institutions and in other countries. To exemplify this, he compared the university entrance examinations administered in Turkey with the international high stakes exams like TOEFL and IELTS and stated that the later ones were more valid.

Furthermore, the participants were also asked whether they knew about any types of validity or not. Although referring to types of validity like content, face, and construct validity is an old paradigm, it is still the common view in Turkey. However, no participant could refer to any type of it.

Practicality, as different from the previous ones, was the term about which all the teacher trainees had a guess. However, some of these guesses were unrelated to the term’s actual meaning in the field of assessment. For example, P1-1 guessed that the instructions given in the assessment tool, which is administered either face to face or online, should be clear enough for the students not to ask questions of the teacher during the assessment. P1-2 noted that practicality might be used to mention the quality of the assessment. Then, he also elaborated that quality meant the success of the assessment, and this success involved the assessment assessing the teaching method or what was taught in the lesson. As different from this view, P1-3 defined practicality as the variety of the assessment tools, and she stated that if a teacher always employed the same type of questions, say open-ended ones, it would threaten practicality. According to P2-2, practicality is related to the learners practicing their language skills in the assessments they take. For example, they should be allowed to write or speak by making sentences in order to answer the questions, and they can read and comprehend the texts and give answers to the comprehension questions. The guess made by P2-3 was based on the meaning of the word “practicality” itself, and she commented, “It was that the assessment tool was appropriate for its purpose both when it was prepared and when it was applied.”

Washback was also another concept that the members of the first case did not know. Regarding the washback, P1-1 expressed that it reminded him of the future effects of the inaccurate teaching when he considered it in terms of education in general. However, he pointed out that he had no idea what the term meant in the field of language assessment. Likewise, P2-1, P2-2, P2-3, and P2-4 expressed that they did not have any knowledge or comments about the concept. On the other hand, P1-4 and P2-5 discussed it in terms of feedback and stated that it referred to the effects of feedback on the students' learning. As different from these, P1-2 briefly defined the washback as the assessment assessing itself. That is to say, P1-2 had the idea that the assessment gave results not only about the students' knowledge or skills but also about the quality of itself. According to P1-3, washback means that the assessment gives information to the teacher about what and how much the students learned.

Strong Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)

In this section, “strong literacy” means the situation in which an individual gives definitions about the key terms that are to a great extent appropriate to the common definition in the literature of assessment. This level of literacy is accepted as one of the necessities for calling a teacher assessment literate. In this vein, some teacher candidates gave accurate definitions or explanations that could be classified under the category of “strong literacy”. Test, assessment, measurement, evaluation, and practicality are accurately defined concepts.

Test was accurately defined by P2-2 and P2-3. Of these, P2-2 stated that it is a tool requiring the learners to use their language skills and that it is generally administered in a written way. Similarly, P2-3 expressed that the test is “a document or a tool with questions and criteria based on measuring a quality or a skill.” She also added that all types of questions could be included.

Five junior PSETs could also make appropriate definitions or explanations about the concept of assessment. For example, three of them emphasized the assessment of language skills, stating that assessment is a process of gathering information about how well students can use their language skills. P1-1 also mentioned that the assessment involves the use of exams and assignments in order to measure the students' knowledge and how they can use their

language skills. P2-2 also referred to the use of language skills and defined assessment as a process-oriented practice in which assignments and tasks are employed.

To test what the individual has learned, it is a study on how much the individual has learned or whether he or she can use the information he has learned. You are doing this together with the individual; I can think of exams, homework, and such things as assessment (P1-1, Semi-structured Interview III).

Measurement was accurately explained by two PSETs, P1-4 and P2-5. P1-4 guessed that measurement means the process of assigning scores at the end of the assessment process. Likewise, P2-5 remarked that this term refers to the quantitative side of the assessment without interpretation and analysis. For explaining this, he said: “A child got this score, but it could only be in a numerical sense without knowing why the child got that score that way, without knowing what was missing.”

Three definitions made about the term “evaluation” focused on the interpretation of assessment results. P2-1 defined it as “the interpretation of the results obtained in the measurement.” Moreover, he gave the example that evaluation could be carried out about the effectiveness of teaching methods and what could be done to improve teaching after measuring the students’ knowledge about the tenses. P2-2 expressed that he could define “evaluation as the grading or interpretation of the measurement result against a certain criterion.” P2-3 commented that evaluation involves decision-making.

Practicality was accurately defined by the teacher candidates by referring to various features such as feasibility, saving time, ease of scoring, and affordability of the assessment tools. For example, P1-4 stated that an assessment method would be practical if it could be used to assess different groups of learners. Moreover, she added that “practicality” means an assessment tool allowing the teacher to score without spending too much time and effort by giving multiple-choice questions as an example.

I don't approve, but I'm going to talk about something that is usually done. You know, it involves a text, and five questions about it. Those five questions are already a test where students will mark something. This is practicality. That is a system that saves time for the teacher and is useful, and s/he clearly collects whatever s/he has received and gives points. This is an example of practicality (Semi-structured Interview III).

P2-1 also remarked that an assessment method or tool should be economical and feasible without any significant limitations. Like P1-4, P2-4 also took feasibility and saving time into consideration when explaining the practicality. In line with this idea, she gave the example that asking the learners to talk about a story after reading it to assess their speaking skills would not be feasible and would take too much time.

Knowledge about the Types of Assessment

Assessment is divided into different types in the fields of both educational and language assessment. Each of these types has unique features, and a language teacher is expected to have adequate knowledge about them. In this regard, the findings of the present research reveal that the junior student teachers generally lack knowledge about the types of assessment. Therefore, their responses consisting of inaccurate or irrelevant comments and explanations, or their lack of any comment at all, constituted the category of poor literacy. Conversely, the explanations and comments that are relevant to the types of assessment concerned constitute the category of strong literacy.

Weak Literacy of the Types of Assessment (First Semi-structured Interview)

The junior PSETs had inaccurate or inadequate knowledge about every type of assessment focused on in this research. Summative assessment is the first of them. Regarding that, three of the participants (P1-2, P2-3, P2-4) said they had never heard of this concept before and could not give any answer. On the other hand, five other participants (P1-1, P1-3, P2-1, P2-2, P2-5) gave an explanation that was not relevant to this type of evaluation. For example, P1-1 defined it as the student summarizing the subject taught to him/her and gave the following explanation:

For example, a subject was explained to the student in a week. The simplest way for this student to convey to me how much, what, how, and how much s/he understands this complex subject in the simplest way possible. And I considered the student a teacher. For example, I taught a topic to a student for a month, covered tenses, and I explained the simple present tense. How can the child use what s/he has learned? How can s/he explain it to me in the simplest way? This is a summative assessment and evaluation from my point of view. [...] It would be for him to explain the whole subject to me in the simplest and most concise way (Semi-structured Interview I).

Formative assessment was also another type of assessment that the participants were not knowledgeable about. That is to say, five of them asked the researcher to skip it without providing any definition. In contrast, P2-1, P2-3, and P2-5 made definitions that differed from each other and did not accurately describe the related type of assessment. Among them, P2-1 described formative assessment as a type of assessment used to guide the thoughts of students, who had different ideas, into a specific pattern.

Table 8

Knowledge about the Types of Assessment

Categories	Codes (Semi-structured Interview I)	Codes (Semi-structured Interview III)
WEAK LITERACY	<i>Summative Assessment</i>	<i>Summative Assessment</i>
	- Unable to define (P1-2, P2-3, P2-4) - Student summarizing the subject taught before in a simple way (P1-1) - Summarizing the assessment (P1-3) - An assessment focusing on the whole subject rather than the specific parts (P2-1) - A type of assessment focusing on a part of a unit (P2-2) - Determining the level of students (P2-5)	- Unable to define (P1-2, P2-2) - Testing what has been taught every week, or every 2 weeks (P1-3) - Giving general information about students' knowledge (P1-4) - Assessing the key points of a unit at the same time (P2-1) - Assessment giving information about language skills without details (P2-5)
	<i>Formative Assessment</i>	<i>Formative Assessment</i>
	- Unable to define (P1-1, P1-3, P1-4, P2-2, P2-4) - A way of evaluating people who go beyond certain boundaries (P2-1) - An assessment type contributing to the student during the assessment (P2-3) - Using tests in line with students' language learning purposes (P2-5)	- Unable to define (P1-1, P1-3, P2-2) - Leading the students to use a specific structure or make a type of sentence (e.g., Using "will future", making long sentences while speaking) (P2-1) - Causing the student to prepare according to the type of assessment (P2-5)
	<i>Diagnostic Assessment</i>	<i>Diagnostic Assessment</i>
	- Unable to define (P1-3, P2-4) - Exams for measuring the student's skill (e.g., 4 language skills) (P1-1) - Defining the student with a definite label (Know or don't know) (P1-4) - A form of assessment that clearly identifies what the student is missing about the subject s/he is directly learning at that moment (P2-2) - An assessment made to obtain a result (P2-3) - Adopting a more scientific approach (P2-5) - More comprehensive than the summative assessment (P2-5)	- Unable to define (P1-3) - Categorization of the scores with some labels (e.g., A, B) (P1-4) - Diagnosing whether the student could use a very specific structure or skill in a very detailed way (Detecting correct or incorrect usage) (P2-1) - An assessment type aiming to detect the student's incomplete learning (P2-2) - Detecting the students' strengths and weaknesses first and then assessing them (P2-4) - Giving detailed results about each language skill or component (P2-5)
	<i>Performance-based Assessment</i>	<i>Performance-based Assessment</i>
	- Unable to define (P1-3) - Evaluation of the participation, behaviors, and attitude during the lesson (P2-4)	- What student does actively in the classroom setting (P1-4) - Assessment focusing on the use of language skills rather than language components: grammar, vocabulary, and pronunciation (P2-1) - Unable to define (P2-2)
		<i>Dynamic Assessment</i>
		- Unable to define (P1-2, P1-3, P2-2, P2-3) - Assessment done during a conversation or a lesson (P1-1) - Continuous assessment (P1-4) - Spontaneous assessment (without allowing the student to get prepared) (P2-1) - Assessment done immediately after teaching (P2-4) - Assessment changing according to place, individuals, time, and topic (P2-5)

<i>Dynamic Assessment</i> • Unable to define (P1-2, P1-3, P2-3) • Spontaneous assessment (e.g., oral exams, interviews) (P1-1) • Continuous assessment (P1-4) • Assessing all activities of the student (e.g., participation, presentation) (P1-4) • An assessment requiring the student to speak immediately about a topic without any preparation (P2-1) • Assessment varying according to the subject (P2-2) • Teacher asking a question during the lesson and giving scores according to the accuracy of the answers (P2-4) • Assessment administered according to factors like learner, teacher, learning environment, and personal characteristics (P2-5) <i>Traditional Assessment</i> • Open-ended questions (P1-4, P2-2, P2-3) • Assessment of the subjects in the coursebook (P1-1) • Performance assignments (P2-3) <i>Alternative Assessment</i> • Not having any knowledge (P1-3, P2-4) <i>Criterion-referenced Assessment</i> • Unable to define (P1-3, P2-3) • Testing the language skills by considering some criteria (P1-1) • Scoring according to how much the student could meet the criterion (P1-2) • Scoring according to specific criteria (e.g., use of conjunctions, 5 points) (P1-4) • A method for measuring exactly the expected knowledge according to certain criteria (P2-1) • Having a specific criterion like a requested skill or a piece of knowledge for being successful (e.g., Using language communicatively) (P2-2) • Related to the formative assessment (P2-4) • Teacher testing according to some criteria (P2-4) • A test on the objectives that should be attained (P2-5) • Testing about what the student has learned or gained (P2-5) • Adopting a more scientific approach (P2-5) <i>Norm-referenced Assessment</i>	<i>Alternative Assessment</i> • Unable to define (P1-1) <i>Criterion-referenced Assessment</i> • Exams for testing a specific subject (P1-1) • Testing done according to a set of criteria (P1-2) • Determining the criteria before assessment (e.g., word count, font) (P1-3) • Items in the rubric (P1-4) • Assessing according to specific criteria (e.g., Criterion: Using simple past tense or could) (P2-1) • An assessment focusing on a specific skill or knowledge (P2-2) • Evaluation of the student's performance according to some specific criteria (P2-3) • Testing according to some specific criteria (P2-4) • Related to the objectives in the curriculum (P2-5) <i>Norm-referenced Assessment</i> • Unable to define (P2-1, P2-3) • Dichotomous scoring (e.g., multiple-choice tests) (P1-2, P2-4) • Evaluating the product as a whole (not focusing on an aspect) (P1-1) • Testing without determining any criteria (P1-3) • Rule (What is forbidden to do or What must be done) (P1-4) • An assessment assessing the subject after teaching (P2-2) • Testing for the acquisition of morally desired behaviors (P2-5)

	• Not having any knowledge (P1-3, P2-2, P2-3, P2-4) • Testing methods focusing on the whole subject without criteria (P1-1) • Dichotomous scoring (e.g., multiple-choice tests) (P1-2) • Involving specific rules (P1-4) • Having a tougher style (P1-4) • Requiring the students to produce (P2-1) • Giving the students more freedom than the criterion-referenced assessment (P2-1) • Testing on moral or national values (P2-5) • Approaching the testing from a moral, cultural, or religious perspective (P2-5)	
STRONG LITERACY	<i>Summative Assessment</i> • Assessing what has been learned so far with one assessment (P1-4) <i>Formative Assessment</i> • Aiming to improve the quality of education after assessment (P1-2) <i>Diagnostic Assessment</i> • Assessments to detect the students' strengths and weaknesses (P1-1, P1-2, P2-1) <i>Performance-based Assessment</i> • Assessing the use of language skills (P1-1, P1-2, P1-4, P2-1, P2-2, P2-3, P2-5) <i>Traditional Assessment</i> • Paper-based examination (P1-1, P1-2, P1-3, P1-4, P2-1, P2-2, P2-3, P2-4) • Multiple-choice questions (P1-4, P2-1, P2-2, P2-3) • Oral exam (Measuring recall or knowledge) (P1-2, P2-4, P2-5) • Fill-in-the-blanks (P1-4, P2-1) • Involving grading (P2-4) • High-stakes exams (e.g., FLT and FLPT) (P2-5) <i>Alternative Assessment</i> • Making a presentation (P2-1, P2-2, P2-3) • Debate (P2-1, P2-5) • Interviews (P1-1, P2-2) • Students' reading and commenting about a piece of news (P1-1)	<i>Summative Assessment</i> • Exams administered at the end of the year or after covering a subject (P1-1, P2-3) • Multiple-choice questions, paper-based exams, and midterms (Previous methods) (P2-4) • No effort to treat the students' errors (P2-4) <i>Formative Assessment</i> • Involving the methods to detect and improve the students' weaknesses (P1-4, P2-3, P2-4) • Aiming to improve the quality of education after assessment (P1-2) <i>Diagnostic Assessment</i> • Assessment to determine the students' strengths and weaknesses (P1-1, P1-2, P2-3) <i>Performance-based Assessment</i> • Assessing written performance (P1-1, P1-3) • Assessing oral assignments (P1-1, P1-3) • Measuring how effectively students can use knowledge (P1-2) • The teacher's observation of the students' in-class performance (P2-3) • The students being more active (e.g., Speaking and Writing) (P2-4) • Asking the student to perform the target skill (P2-5) <i>Traditional Assessment</i> • Multiple-choice questions (P1-1, P1-2, P1-3, P1-4, P2-1, P2-2, P2-3, P2-4) • Fill-in-the-blanks (P1-1, P1-3, P1-4, P2-1) • Oral exams (P1-1, P2-3, P2-5) • Open-ended questions (P1-3, P1-4) • Assignments (P1-1) • Quizzes (P2-3) • True-False statements (P2-4)

• Process-oriented assessment (P1-2) • Focusing on the student's development (e.g., writing skills) (P1-2) • Collecting the student's products (P1-2) • Group assignments (P1-4) • Focusing on performance (P1-4) • Performing a theater play with 5 or 6 students (P1-4) • Introducing your city (P1-4) • Writing a story about a song (P1-4) • Group talk (P2-1) • Question-and-answer (P2-1) • Involving games, visuals, and music (P2-1) • Taking the context of the exam outside the paper (P2-2) • Performance-based (P2-3) • Aiming to reveal the language skills (P2-3) • Listening to and answering open-ended questions (P2-3) • Portfolio (P2-3) • Compatible with the opportunities of the 21st century (P2-5)	<i>Alternative Assessment</i> • Making a presentation (P1-2, P1-3, P2-1, P2-2, P2-4) • Assignment (P1-2, P1-4) • Debate (P1-3, P1-4) • Process-oriented assessment (P1-2, P2-5) • Role-playing (P2-2, P2-3) • Group talk (P2-1) • Video and audio recording (P2-2) • Group work (P2-2) • Portfolio (P2-3) • Peer-assessment (P2-3) • Self-assessment (P2-3) • Asking the students to perform (P2-4) • Assessing with games (P2-4) • Assessment done with different tools (P2-5)

P2-3 asked, "Is formative assessment a type of assessment that contributes to the student in the assessment?" and commented, "It could be an assessment that can shape that student even during the assessment." P2-5, dealing with this concept from a different perspective, explained it as assessments made in school or in private courses in accordance with the purpose of the student's language learning. For example, if a student needs to score well on a high-stakes exam with multiple-choice questions, administering multiple-choice questions to the student is formative assessment, or if the goal is to learn to use language communicatively, administering communicative assessments aligned to that goal is formative assessment.

The student teachers in Case 1 also do not know what diagnostic assessment is. In this regard, two out of nine respondents could not give any answers, while the other seven did not give the correct answer. Of these answers, for example, the answer given by P2-1 includes that the diagnostic assessment aims to identify what the student does not know more than s/he does. P2-3, who stated that he had never encountered this type of assessment before, stated that diagnostic assessment is "an assessment aimed at drawing a conclusion" based on the word diagnostic, likening it to the methods used in hospitals to make a diagnosis. It should be noted that although these responses seem to be partly related to diagnostic assessment, they are in the category of weak literacy because they do not mention the most important features of diagnostic assessment, which are that inferences are made about the teaching process after the assessment and that this teaching process is designed according to these inferences.

The teacher candidates usually have knowledge about performance-based assessment, while dynamic assessment is a type of assessment that none of them have knowledge of. Although no participant has heard of this type of assessment before, some of them have tried to guess what it is. Some of these estimates are that it is an assessment done without any preparation, that includes all kinds of activities that students do in the educational-teaching process, and that it is an assessment applied taking into account different factors.

Participants were generally aware of traditional and alternative assessments, while some mentioned practices that did not fully match these types of assessments or did not make guesses or comments. For example, when they explained the traditional assessment, they stated that it included open-ended questions and performance tasks and that they focused on

the topics in the textbook. When it came to the alternative assessment, P1-3 and P2-4 stated that they had no knowledge about it.

CRA and NRA are two types of assessments that the participants were not familiar with. Although some participants stated that criteria were used in CRA based on the name of the type of assessment, they were included in the category of poor literacy because they did not emphasize making decisions about the success of the students by comparing their performances with these criteria rather than comparing their performances with the performances of other students in the group. As for NRA, 4 participants did not have any predictions or comments about it, while the others linked it to a total focus on a comprehensive topic, objective marking, moral values, having specific rules, or giving students more opportunities for production. P2-5 compares these two types of assessments as follows:

CRA is a test on which achievements should be gained while teaching that lesson to the student. When I think of norms, I think of testing moral, traditional, or national values. The differences between them are that I think CRA is more about what the student has learned or gained. It is a bit more scientific in this way. The NRA is a bit more like this, you know? It can be approached morally, culturally, and religiously (Semi-structured Interview I).

As the excerpt above shows, the participants sometimes made comments based on the names of assessment types.

Strong Literacy of the Types of Assessment (First Semi-structured Interview)

Some respondents made relevant comments or predictions in relation to some types of assessments. Therefore, these responses constitute the high literacy category. Only one participant gave correct answers about summative assessment, while some others gave correct answers about performance-based assessment, traditional assessment, and alternative assessment. To explain in more detail, P1-4 guessed by emphasizing that the summative assessment was a one-shot assessment rather than a process and stated as follows:

There could be one assessment at one time, at one stage, I guess. It seems to me that this is something like that. How can I put it? To evaluate a student at one time with a single written paper or, for example, something oral. That's how it seems to me. It's like being able to ask and evaluate whatever they have learned until that day (Semi-structured Interview I).

Regarding performance-based assessment, seven participants stated that this assessment requires students to use their language skills. They also emphasized that these assessments

involve tasks or assignments such as role-playing, reading and responding, communicating, and writing a critique about a work. In this vein, P2-1 made such an explanation about the performance-based assessment:

This is more like productive skills. More for the use of skills. Apart from the things we already know, such as multiple choice, fill-in-the-blanks, and so on, the student can create and produce something. It seems like an evaluation method for skills such as presenting or talking about the subject, reacting to it, or writing about it (Semi-structured Interview I).

As to traditional and alternative assessments, student teachers know what these assessments are. In particular, they are aware that traditional assessment methods, which they have experienced for years, consist of paper-and-pencil tests, short-answer questions such as multiple-choice and fill-in-the-blank, and the existence of scoring or grading. Similarly, although they have not had much experience before, they know that alternative assessment methods consist of methods such as making presentations, discussions, interviews, writing stories, and preparing portfolios, which require the use of language skills.

Weak Literacy of the Types of Assessment (Third Semi-structured Interview)

The findings of this research revealed that the members of the first case were also not knowledgeable enough about the types of assessment. CRA and NRA refer to the two families of assessments that have different functions (Brown, 2005, p. 1). CRA, which involves making a judgment according to a predetermined criterion (Douglas, 2010, p. 68), was one of them, and the participants made different remarks on what it is. P1-1 accentuated that the exams administered for CRA focus on a specific subject or ability, and the test maker has an expectation to get information about the student's knowledge or performance on this subject or ability. According to P1-2, P2-1, P2-3, and P2-4, this is the type of assessment according to a set of criteria, like using the simple past tense or modals accurately. On the other hand, P1-3 explained the CRA as determining the criteria before the assessment. Then, she gave examples for the criteria, such as the word count, font size, and number of paragraphs for the assessment of writing skills. Like P1-2, P1-4 also emphasized the rubric while explaining the CRA. In this regard, she added that this refers to the situation where the scores are assigned according to whether the criteria are met or not. Unlike the others, P2-2 defined CRA as "assessments that measure a particular skill or knowledge by considering

only that knowledge.” P2-5 mentioned the objectives in the curriculum and related CRA to what was expected of the learners.

NRA is the second family, which involves “classroom-level decisions, such as diagnostic and achievement decisions” (Brown, 2005, p. 1). As it was for the CRA, the participants could also not accurately explain this sort of assessment, although they had various comments on it. For instance, P1-1 made the guess that the assessor takes the whole product of the student into consideration rather than focusing on a specific subject as in the CRA. On the other hand, P1-2 expressed that the NRA involves dichotomous and absolute scoring, as done in the multiple-choice tests. As for the answers by P1-3 and P1-4, they also had different views. P1-3 guessed that the NRA means the testing without determining any criteria beforehand. In contrast, P1-4 stated that the word “norm” reminds her of strict rules, and the CRA involves these rules that the students should obey throughout the assessment process.

As for the second graders, P2-1 and P2-3 could not provide any definition of the CRA. Moreover, P2-1 pointed out that he could not differentiate it from the NRA. On the other hand, P2-2 referred to the NRA as an assessment administered on the subject covered in the lesson and gave the example that the usage of simple present tense was assessed with a test after it was taught by the teacher. The NRA was also defined by P2-4 as a test consisting of multiple-choice questions and having predetermined rules like 4 points for each correct answer. Based on the word “norm”, P2-5 commented that the NRA was testing whether the learners acquired the morally desired behaviors such as kindness and love for homeland and nation.

Furthermore, various approaches to assessment were the focus of this research. In this vein, the participants were asked to explain these approaches. Traditional and alternative assessments were two of them, and it was found that they had generally enough knowledge about them. However, only P1-1 could not provide an accurate explanation about alternative assessment.

In contrast to the findings regarding traditional and alternative assessments, the research revealed that these pre-service teachers are not adequately knowledgeable about summative and formative assessments. While P1-2 and P2-2 could not make any predictions about the summative assessment, P1-3 articulated that the summative assessment involves testing how

much the students have learned each week or every two weeks before the actual exam. P1-4 commented that this sort of assessment gives general information about the students' knowledge or abilities by focusing on the main points. Additionally, P2-1 made an incorrect definition and explained summative assessment as a procedure of simultaneously assessing the basic points in a unit. P2-5, who considered summative assessment a general assessment of language skills, stated that this assessment does not provide detailed information about language skills.

Summative assessment is a bit more general, in the form of summarizing the general skills of the student in general. For example, how much can s/he speak? Good, average. But it is not known at what level. Good or average. Like this. As the name suggests, it is an assessment that summarizes and does not go into more detail (Semi-structured Interview III).

Formative assessment is another type about which the junior teacher candidates did not have enough information. Three pre-service teachers, P1-1, P1-3, and P2-2, could not define it and stated they did not have any information about the formative assessment. P1-2 emphasized the name "formative assessment" and pointed out that it was an assessment type aiming only "to shape our way of teaching" and "improve the quality of it". However, he could not give any explanation about its role and effects on learning. P2-1 was not also able to provide a correct answer and explained the formative assessment as an assessment requiring the learners to make a certain length of sentences by using some specific or predetermined structures like "will future." The answer given by P2-5 was that the formative assessment was a case of the learners studying or learning the language according to the types of questions used in the assessment.

The junior PSETs also have problems defining or explaining the diagnostic assessment. Of these, P1-4 expressed that the diagnostic assessment consisted of the categorization of scores.

It feels like this. I don't know for sure, but. And what kind of result is in front of us? There are evaluations, and should I say the label of this evaluation? For example, 75 has such a label, but 90 is like this. I mean, I have this thing of diagnosis, of putting something into a mold. [Do you mean something like categorization?] I mean exactly. I mean, for example, we say, I bought A or B. For example, I feel like this is something like that. It seems to correspond to it (Semi-structured Interview III).

According to P2-1, the diagnostic assessment involves detecting in detail whether the student is able to use a very specific structure or subskill. In line with this definition, he says, "it is like an assessment for pinpoint detection." The aim here is to unravel how correctly or incorrectly the learner is using the grammar structure or the language skill in question. As

another similar explanation, P2-2 guessed that the purpose of diagnostic assessment was “to reveal precisely which aspects of certain knowledge a student had learned incompletely.”

The guess by P2-4 is that the diagnostic assessment is the teacher first identifying the learners’ strengths and weaknesses before the assessment and then carrying out the assessment according to these detections.

Since I go directly from the diagnosis, is it possible to define what the students are lacking or better at and to apply certain assessments to them? But, well, of course, this identification does not happen at the assessment stage. It's like before. [Do we make a definition beforehand and then apply an assessment and evaluation?] Yes, I guess because it is diagnostic, because I think it is like putting certain diagnoses directly beforehand and then providing students with measurement and evaluation for them (Semi-structured Interview III).

P2-5 expressed that the diagnostic assessment might be “a bit more analysis-oriented” and clarified it with the following sentences:

Diagnostic, I think, is a little more analysis-oriented. In other words, four basic skills are separated in the exam. At the speaking level, sentence use is good and grammar use is bad, for example. At the writing level, the student has internalized words and patterns. But, for example, s/he does not know the rules of spelling. In listening, the student has a general ability to understand, can understand, but has problems when it comes to details. There can be more detailed evaluations (Semi-structured Interview III).

Some of the junior PSETs could also not give accurate definitions or explanations regarding the performance-based assessment. Of these, P1-4 and P2-3 referred to the learners’ participation in the lesson as “performance”. In line with this, P1-4 stated this is not a paper-based assessment and involves the students’ physical and emotional involvement as follows.

I mean, I don't know if a bad performance can be seen as a performance in an exam. I think it's more of a performance, an assessment in which the person is active with his/her own body and soul. You know, an assessment that is not about the information on a paper but about what he/she does in the classroom environment in general, or even in almost the whole classroom environment. As we say, performance, it is like that kind of thing; it is a type of assessment that is a little bit broader and more active (Semi-structured Interview III).

According to P2-1, the language components are not assessed in the performance-based assessment. Rather, this involves methods focusing on the assessment of the use of language skills. On the other hand, P2-2 could not give any definition or explanation.

Dynamic assessment is the last type of assessment about which the teacher candidates do not have any knowledge or cannot provide an accurate explanation. To be more precise, four PSETs, P1-2, P1-3, P2-2, and P2-3, were not able to define dynamic assessment. P1-1 made the guess that “dynamic assessment can be defined not in the context of an exam but in the context of a conversation between a teacher and a student about a memory or something said in a lesson.” While P1-4 states that dynamic assessment is a state of constant measurement

and evaluation, P2-1 calls it spontaneous assessment in which the students are not allowed to get prepared for the assessment. According to P2-4, it is the assessment that is carried out with some specific questions asked immediately after the instruction in order to reinforce the taught content. Moreover, the learners get feedback according to their answers to the questions asked by the teacher. As different from these explanations, P2-5 expressed that the term “dynamic assessment” means an assessment changing according to the place and time in which it is conducted, the individuals taking it, and the topic on which it focuses.

Strong Literacy of the Types of Assessment (Third Semi-structured Interview)

In contrast to these indicators of weak LAL, there are also accurate explanations and definitions pointed out by the junior teacher candidates. For example, all of them could accurately explain the traditional assessment either by giving examples or mentioning the nature of it. Multiple-choice questions, fill-in-the-blanks, and open-ended questions were the items the PSETs frequently mentioned regarding this type of assessment. Moreover, how the traditional assessment is carried out is also addressed. In this vein, they expressed that it could be administered as paper-based or oral exams.

Traditional assessment is usually a type of assessment that subjects the student to a specific written exam and then uses a score as a result of that written exam. Usually, these 40-question tests or direct written exams that include fill-in-the-blank questions and put them in front of us with test questions (Semi-structured Interview III).

As with the traditional assessment, the PSETs explained the alternative assessment with examples and statements about its nature. Role-play, presentation, debate, games, self- and peer-assessment, and portfolio are the methods given by the pre-service teachers as examples of alternative assessment. For example, P2-2 made the explanation as follows:

Alternative assessments, on the other hand, are those that measure the student's performance through products that the student creates. For example, video recordings, audio recordings, live presentations, student presentations, or group work, role-plays. These can be included (Semi-structured Interview III).

Moreover, they mentioned the features of the alternative assessment. That is to say, they stated that the alternative assessment is process-oriented, focuses on students' performance in using language skills, and appeals to students' creativity.

As for the summative assessment, it was correctly explained by only one pre-service teacher, P1-1. His answer focused on when it can be carried out and what content can be assessed in it.

Maybe end-of-year exams and similar things. I can think of something like that. Putting a summary of all they have learned in 1 exam or 1 homework without detailing it, I think this would be in order to be able to see all they have learned in 1 summary [...] Of course, it does not have to be at the end of the semester. For example, if a subject has been worked on, I think it would be a summative assessment if at the end of the semester, instead of dealing with it piece by piece, the person uses the whole subject and evaluates the summary of how much he/she has actually learned (Semi-structured Interview III).

Formative assessment was also accurately explained or guessed by the three PSETs. All of them emphasized that it involves efforts to improve the weaknesses of the students by detecting them through assessments. Of these, while P1-4 referred to it as “an incentive or shaping on how to complete the missing aspects,” P2-3 expressed that it was an assessment focusing especially on quality and “not only identifying the students’ weaknesses but also compensating them.” Likewise, P2-4 pointed out that it was “an assessment method in which methods are used to correct the students’ shortcomings or correct their mistakes.”

Two first-year PSETs’ answers about the diagnostic assessment were also correct and classified under the category of “strong literacy.” P1-1 said: “An exam that tests all four abilities is actually a diagnostic assessment in my opinion. Because it is something that is done to understand what that individual is inclined towards and similar things.” From another perspective, P2-3 accentuated that the diagnostic assessment focuses on how much prior knowledge the students have about a subject area.

This is the type of assessment that aims to make a diagnosis. That is, if we think of it as a diagnosis, it is the type of assessment that aims to see how much the student knows. It is the first assessment. Probably the first assessment is to see how much the student knows. It is an assessment, a measurement made for the purpose of diagnosing (Semi-structured Interview III).

Performance-based assessment was accurately explained by two junior PSETs. Both of them referred to the use of language skills like speaking and writing for addressing performance. According to P1-2, the performance-based assessment pays attention to “the part we call effective use of how successful our students are in practice” or “how s/he uses knowledge.” In line with this, P1-3 also answered that how the students were able to perform their language skills would be at the center of this assessment by focusing on “how they wrote” and “how they spoke.”

Awareness of the Factors Affecting the Quality of Language Assessment

This section includes the findings of the teacher candidates' awareness regarding the factors having the potential to affect the language assessment. These factors consist of the assessment experiences they had, the teacher's qualities to conduct quality assessment, and the language learning and teaching approach.

Awareness of the Factors Affecting the Quality of Language Assessment

STRONG LITERACY

The Effects of Assessment Experiences on Language Assessment Approach

- Leading the teacher candidate to notice what not to do (P1-1, P1-3, P1-4, P2-1, P2-2, P2-3, P2-4)
- Using the appropriate methods used by the teachers (P1-2, P1-3)
- Leading to use classical assessment methods (Negative effects) (P2-1)
- Demotivate the student (P2-2)
- Learning about ideal assessment (P2-2)

The Necessary Teacher Qualities for Quality Assessments

- The teacher being fair (P1-1, P1-2, P2-2)
- Teacher's knowledge about the field of language education or teaching (P1-4, P2-1)
- Being competent in choosing, preparing, and administering quality language assessments (P2-1, P2-3)
- Teacher having knowledge about the assessment topic (P1-2)
- Teacher's ability to interpret the student's performance (P1-2)
- Teacher's efforts to improve himself or herself in terms of professional competence (P1-4)
- Teacher being aware of the concepts of WE, ELF, and ESL (P2-1)
- Teacher being competent in teaching and assessing by considering WE, ELF, and ESL (P2-1)
- Being able to determine limits for language skills according to the student's language level (P2-1)
- Being able to monitor the changes in the students' language development (P2-1)
- Having enough knowledge about the formal and informal expressions (P2-2)
- Being creative (P2-2)
- Loving his or her job (P2-2)
- Knowing what the concept of "language skills" encompasses and requires (P2-4)

Teacher's Competency in Language Skills

- Teachers being competent in four language skills for conducting quality assessments (P1-2, P2-2, P2-3, P2-4, P2-5)
- Teachers having knowledge about various uses of language (P2-2)

Theoretical Knowledge about Language Assessment

- The importance of theoretical knowledge: A guide for assessment (P1-4, P2-1, P2-2, P2-3, P2-4)
- Being knowledgeable about which techniques are appropriate to assess which skills (P2-1, P2-2, P2-3)
- Teacher's having enough knowledge about the assessment method to be used (P1-2, P1-4)
- Being knowledgeable about how to do the assessment and getting results (P2-3)
- Being knowledgeable about the factors affecting the quality of assessment (P2-3)
- Teacher's knowing that the assessment content should not be strictly similar to the lesson content (P2-4)

The Effects of Language Teaching Approach on Language Assessment Practices

- Consistency between the teaching and assessment methods (P1-1, P1-2, P2-1, P2-2, P2-3, P2-4)
 - Inconsistency between the teaching and assessment content (P1-4, P1-3)
 - Teaching grammar and vocabulary, and assessing them (P2-1)
-

The findings reveal that the participants have awareness of how their assessment experiences affect their approach to language assessment. Therefore, their answers fall into the category of strong literacy. As the most common code, they mentioned that the traditional, mechanical or non-communicative assessment techniques or tools they experienced in the English lessons focused on only their reading skills or grammar and vocabulary and thus caused them

to ignore other language skills. However, as the training they took in the ELT department at university required the communicative use of all language skills, these participants noticed that these assessment methods, which they had previously experienced, were not much appropriate for the language assessments. To illustrate this, P2-1 stated the following words:

I agree with my other friends. Until university, I also saw different types of assessment in high school, but my awareness was low at that time. As I saw various course-specific assessments in various courses in addition to the courses I saw at the university, I realised that different assessments could be made in conjunction with the content of the course, that a different method could be followed for each ability, and perhaps even that this could be done for students, if the number, time or different limitations allow it. As my other friends said, these assessments were always something until university: they were harmful, I should say bad. It put me into the normal, usual exam mould. If you had put me in a classroom with the English I knew when I came out of high school, I probably would have used the test method or fill-in-the-blank methods. But now, at least in the 2nd grade, I realise that I have to use different assessment methods and that they can be classified and changed according to students, classes and regions (Focus Group Interview I).

Furthermore, two of the freshmen PSETs reported that they would also employ the assessment methods that they thought were appropriate. For example, P1-2 said that the assessment experiences had a big impact on language learners' approaches to assessment and they tended to imitate the assessment tools used by their teachers. However, he mentioned the importance of creative thinking for today's conditions and stated that pre-service teachers started to acquire an approach towards making more communicative evaluations by using their creativity. For this reason, he emphasised that he would not only use the same assessment methods used by his teacher but also make adaptations on these assessment methods.

As for the qualities that an English teacher should have to conduct quality assessments, the student teachers referred to various features indicating that they had a high level of awareness. To clarify, teacher being competent in the four language skills is the oft-cited quality in addition to the teacher having theoretical knowledge about language assessment. Regarding the teacher's proficiency in using the language skills, the teacher candidates emphasized that this was a prerequisite for quality assessments. For example, P1-2 wrote the following sentences in his journal:

"A teacher who is able to assess language skills should have a very good knowledge of language skills. It would not be healthy for a teacher to make an assessment in a subject in which s/he is not well qualified" (Opinion Journal II).

Similarly, the teacher candidates mentioned the teacher's theoretical knowledge of language assessment as another important factor for conducting quality assessments. This includes knowledge about the assessment methods to be used, what to include or not to include in the

assessment content, how to obtain the results regarding the target construct, and potential factors that will influence the quality of the assessment. Of these, the knowledge about the assessment methods is the most frequently mentioned knowledge area and it refers to the teacher's being knowledgeable about which assessment methods or techniques are appropriate for assessing which language skills and components and their practicality. The determination of the content of the assessment mentioned by P2-4 involves the teacher's having the knowledge that the assessment should include related but new content rather than including the exact same one that the teacher explained in the lesson. Likewise, P2-3 noted that an English teacher should take the objectives in the curriculum into consideration while preparing the assessment and know how to obtain results regarding his or her students' proficiency in the target language skills or components. Moreover, she pointed out that the teacher should have knowledge and awareness about what kind of factors impact the quality of assessments.

In addition to these, according to P2-1, the necessary knowledge that an English teacher should have also include the knowledge about the topic that the assessment is about, the field of language teaching and the concepts of English as a lingua franca (ELF), World Englishes (WE), and English as a second or foreign language (ESL/EFL). Regarding the knowledge about the assessment topic, P1-2 accentuated that the teacher being knowledgeable about the assessment content may help him or her to prepare impartial assessments by avoiding the offensive or advantageous content. P1-4 and P2-2 addressed the relationship between language teaching and assessment and the importance of the teacher's competency regarding the field of language teaching in order to create the conditions for quality assessments. For example, in her Opinion Journal, P1-4 noted the importance of the prospective teacher's professional development as well as the training s/he received at university.

First of all, it should be known that it is not possible to make a qualified and adequate evaluation unless there is a qualified / adequate teacher. In order to have this, not only the institutions and people who are trained, but also the target person (teacher) should be interested in studies, environments, people and activities that will improve and make him/her self-sufficient. To be more precise, if a person who graduates from the faculty of this department only tries to finish the school and fulfil the given tasks, there will not be much tangible accumulation that he will add to himself at the end of the process (Opinion Journal II).

Like P1-4, P2-1 also reported that the teacher should be knowledgeable about the traditional and contemporary teaching methods and techniques of English as a foreign or second language. Additionally, he mentioned that the teacher should be aware of the concepts of ELF, WE, and ESL and regard these while teaching and conducting language assessments.

Implying that the development process of the learner's interlanguage (Selinker, 1972), he explained why this awareness is important for the language assessments in the sentences below:

First of all, [the teacher] needs to be familiar with concepts such as WE, ELF, ESL, etc., i.e. to be aware that each of the students to be assessed is a unique user of English, so as not to create a barrier that is literally impossible or near impossible to overcome when testing or assessing language skills. Otherwise, students will have to be assessed not only within the specific rules of the language appropriate to that skill (sentence structure, pronunciation, speaking in relation to the topic, etc.), but also under certain language uses that do not actually exist within the scope of the assessment. This situation will result in a chain of errors in monitoring the student's developmental process in language acquisition and will also lead to the fact that the results obtained during and at the end of this process do not belong to the student's original achievement. For all these reasons, the teacher should have self-awareness of the mentioned concepts and should be able to reach a certain level of competence in their teaching and assessment (Opinion Journal II).

Awareness of how the language teaching approach that the teacher adopts may impact the language assessment practices that the teacher will carry out is also an important dimension. As the indicators of this awareness, the junior PSETs referred to the negative washback effects of the high-stakes exams like FLPT, the consistency between the teaching and assessment methods, and the in/consistency between the teaching and assessment content. Regarding the first of these, P1-1 and P1-2 emphasised that the high-stakes examinations in Turkey have led to the use of methods in teaching English, especially to 12th graders, in line with the content of these examinations. As a result, the teacher exploits the pen-and-paper exams and multiple-choice questions as compatible with the teaching methods and the high-stakes exams. Similarly, four of the second-year teacher trainees also addressed the consistency between the teaching and assessment methods. However, they did not refer to the washback effects and pointed out that since some English teachers do not have the competence to use communicative language teaching methods, they teach English with traditional teaching methods and accordingly they employ traditional and non-communicative tools for assessment.

Literacy in the Skills Component of LAL

Assessing the Language Skills

This section involves the findings related to how the junior teacher trainees will assess their prospective students' language skills. Each of the four language skills (listening, speaking,

reading, and writing) is separately presented under sub-headings. Furthermore, the findings regarding each skill are given for two groups of learners: young learners and adolescents. In this research, the previous one refers to the students studying at a primary and lower secondary school. On the other hand, the second one is used to refer to the students studying at an upper secondary school. Although the young learners group involves students from various grades and ages, the appropriateness of the answers for the students from the primary or lower secondary school is taken into consideration.

Assessing Listening

The junior PSETs resorted various ways of assessing the listening skills of their future students. While some of them were more traditional and non-communicative techniques, others were more appropriate for the communicative assessment of listening skills. Therefore, the practices that are not in line with the communicative assessment were classified under the category of *weak literacy* while the practices that are congruent with the communicative assessment were classified under the category of *strong literacy*.

Assessing the Listening Skills of Young Language Learners

In this research, the group of young learners consists of students from both the primary and lower secondary schools. Regarding the assessment of the listening skills of these students, some junior PSETs stated they would employ non-communicative, traditional assessment methods such as fill-in-the-blanks, multiple-choice questions, and true-false statements. P1-2 wrote, “I assess their listening skills with a test that includes fill-in-the-blank and multiple-choice questions, like a classical listening test.” Similarly, P1-3 answered that she would conduct discrete assessments on the students’ language skills. In line with this view, she pointed out that her assessments would draw on fill-in-the-blank and multiple-choice items. Moreover, she had the belief that the teacher reading the text aloud was more appropriate for the students’ level than having them listen to an audio recording.

According to the topics we cover in the lesson, we can determine a topic appropriate to the ages of the students. For example, I think that if we read something to young children, it will be good for them. Making them listen to it by playing a CD is not suitable. We should read it ourselves. [When you think about it at the upper levels, do you say that we will listen to them on CD instead of reading them?] Yes. Listening to it on CD. Because if we read it ourselves, it would not be

appropriate for their age. I think it would be better to listen to it on CD (Semi-structured Interview III).

P2-1 also wrote in the scenarios which items he would employ for assessing the listening skills of the young language learners.

For the assessment, I would prepare video content suitable for the topic and ask the students to answer the fill-in-the-blanks, picture word matching (if it is said in the video that breakfast is disliked, match the breakfast picture and the sentence I don't like), and multiple-choice questions by making a connection between the visuals and the speech. I would ask them to read the first part of the questions before the video and watch the video from beginning to end. Then I would ask them to read the remaining two parts, watch the video again, and do these parts (Scenarios for Foreign Language Assessment).

P2-2 and P2-3 also pointed out in the interview that fill-in-the-blanks and multiple-choice questions would be used in their listening assessments. Here, P2-2 emphasizes that the fill-in-the-blank items should be simpler than the multiple-choice questions because he thinks that the earlier one is more open-ended than the later one. Moreover, he adds that “in multiple-choice questions, [he] would ask general questions that directly address the text.” As for P2-3, she stated the fill-in-the-blank items would be adopted in addition to the open-ended questions. With regard to these items, she also accentuated that they would allow the students to make some interpretations about the appropriate answer for the blank by drawing on the given parts of the sentence.

Table 9

Assessing the Listening Skills of Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill in the blanks (P1-2, P1-3, P2-1, P2-2, P2-3) • Multiple-choice questions (P1-2, P1-3, P2-1, P2-2) • Having difficulty assessing young learners (P1-2, P1-4, P2-1) • Student translating the sentences said by the teacher (P1-1) • True-false (P2-1) • Listening exam (P2-5) • Teacher reading aloud text for listening due to students' English level (P2-5) • Paying attention: Percentage of comprehending text and catching words (P2-5)	• Open-ended questions (P1-1, P1-3, P1-4, P2-2, P2-3, P2-4) • Asking the students which words they heard in the audio (P1-4, P2-2, P2-3) • Matching (P2-1, P2-2, P2-4) • Assessing through physical reactions (P1-1, P2-1) <i>Listening and Speaking Integrated</i> • In-class discussion after listening to the assigned story or podcast (P1-4) • Making a short presentation to answer the questions (group work) about the listening text (P2-2) • Asking students questions about using the words in the audio (P2-2) • Answering the questions asked about the audio (P2-4) • Pronouncing the words after listening (P2-4) <i>Listening and Writing Integrated</i> • Asking questions about the theme of the text, the description of the events, or the prominent characters in the text (P2-2) • Asking students to name the activities given in the images (P2-2) • Asking students to write appropriate activities for the appropriate hour (P2-2) • Answering the questions asked about the audio (P2-4)

Furthermore, as an indicator of weak literacy, three of the junior PSETs explained that they had difficulty assessing the young learners' listening skills. In other words, although they mentioned some items for assessing listening, they also noted that they were not confident about assessing this skill of the primary school students for some reasons. For example, the students' low English level was a cause of the insecurity felt by P1-2 and P2-1.

Listening skills for primary school—I don't know if this education is given. I don't know how we can conduct an assessment for listening since the student does not have such a high level of English in primary education. It will be a little challenging for the students (P1-2, Semi-structured Interview III).

On the other hand, P1-4 gave the reason that she had not previously focused on listening skills.

Actually, I have difficulty determining if the student really improved his listening skills at the end. But I don't know. I also don't know how long I can expose that child to listening. I am only sure about engaging him with the very basic things. But about the objectives. Also, it's like I've come to this point by not caring so much about listening. It's a little hard to think about it now. So it's like something I need to think about (P1-4, Semi-structured Interview III).

Furthermore, in the interview, P1-1 answered that he would ask his students to translate what he said to them in English.

During listening, I try to understand how much s/he can visualize the Turkish version of the simple sentences in his/her head. For example, when I say "I have 5 siblings" in English, when I ask this child in Turkish, I want him or her to be able to tell me how many siblings I have in a simple way (P1-1, Semi-structured Interview III).

Some teacher trainees mentioned more communicative and authentic assessment practices in contrast to these indicators of *weak literacy*. Although they generally experienced the assessments done with traditional assessment methods, the participants sometimes drew on their experiences in language classes as young learners of English while referring to these practices. Labeled as indicators of *strong literacy*, these involve various practices like acting according to a song, narrating what the teacher said, answering some questions asked about audio, and discussing the content of audio with peers.

Of the participants, P1-1 believes that the language skills of the young learners should be assessed discretely. Regarding this, he gave the reason that the student's weakness in one skill would negatively affect his performance in another one, and this would demotivate the student during the assessment. When it comes to how to assess listening, he states that he will prefer to use recordings that will be enjoyable and useful for the students in their daily lives. Remembering that he had listened to a song called "Head, Shoulders, Knees, and Toes" for introducing the body parts, he expressed that he would have the students listen to a song

for listening assessment. As for the assessment practice, he explained it in the following excerpt:

I would like them to show themselves. For example, we mentioned that there is a song about body parts. We listened to it when we were little. In one part of the song, for example, the head is mentioned. I would like him to show his head so that when the child shows what is what, he will notice himself. In a practical sense, looking practical seems to be more effective for me in such small age groups. Asking them to show what they've heard in real life. Most of the things that the audio mentions are things that the students can show in their daily lives. They are things like fruits, vegetables, and body parts (Semi-structured Interview III).

Additionally, P2-1 stated that the teacher would give instructions to the young language learners and expect them to respond physically to those instructions. He will also focus on the oral answers of the learners after listening to the sentences said by the teacher.

P1-3 would ask her students what they understood from the audio. Similarly, P1-4 will ask about the topic of the listening track and prefer to “mostly focus on their associative power and ability to connect, rather than expecting them to understand everything they listen to”. P2-2 will employ a listening text that is simple and not too detailed. “The questions that are not very detailed, but rather questions that focus on the general theme of the text”. Moreover, he will focus on “the description of the events in the text being listened to, or if there is a character who is prominent in the events; in short, what the text highlights.” Through these questions, he will aim to assess not only the learners’ understanding of the audio but also their inferences and interpretations. P2-3 expressed and wrote that she would ask short, open-ended questions in a written exam in which she would have her students watch a video or listen to an audio recording.

P2-1 will ask his students to match the sentences with the pictures according to the audio they will listen to. He gives the example of matching the picture of breakfast with the sentence “I don't like” if the video says “I don't like breakfast”. P2-2 wrote in the scenarios that he would give the lower secondary school learners pictures representing specific activities and different time periods. He would ask his students “to name activities first,” and then they would match the activities with the periods. P2-4 noted in the scenarios that the fourth graders would be assessed with matching items as well as sentence-making and comprehension questions instead of true-false and multiple-choice questions.

P1-4 stated that she would conduct an informal assessment and have her students watch an animation or a cartoon and assess their listening skills by stopping at certain places and asking them which words were used. In the scenarios, P2-2 and P2-3 wrote that they would

employ short listening tracks and carry out an assessment that would ask the learners which classroom objects they heard and what their colors were.

In their informal assessments, which integrate the listening and speaking skills of the students, P1-1, P1-3, and P1-4 pointed out that they would ask wh-questions to the students about the topic or content of what they listened to. For the listening material, while P1-3 preferred to read the text herself, P1-1 and P1-4 expressed that they would use songs, animations, and cartoons that would more easily attract the attention of the students. Moreover, it is worth noting that while giving this answer, P1-1 remembered again that his teacher had them listen to the Beatles' Yesterday song and asked them about the topic of the song.

As another practice of integrated assessment, P1-4 accentuated that she would assess the listening through in-class discussions among the students based on the topic mentioned in a song or a podcast. Believing that English should be the only language used in the classroom, she would enable the students to both listen and speak through this practice of assessment. Furthermore, assessing the students in the long run rather than through a one-shot practice is another issue she emphasized, as given in the quotation below.

For example, there are some storytellings or podcasts we listen to. I have the students listen to the podcasts in English at these levels. For example, a 20-minute podcast. Friends, I am very clear about this: the English lesson should progress by speaking English, not Turkish at all. Then, friends, what do you think about it? Do you agree with what he thinks? Or, if you have a different opinion, what are your reasons? It will be the activities that make the whole class active. This is developing with the activities in process. What else could possibly be on the exam? In my opinion, it is not necessary to assess listening with the exam. I would like to proceed with these and reach an evaluation. Because it doesn't make much sense to me to evaluate based on what s/he listens to in a day. That's why I would do that by considering what he did with them in that process, or what he acquired or how far he has progressed (Semi-structured Interview III).

P2-2 will ask his students to make a presentation about the answers given to the questions asked about the audio. The students will create these presentations in groups, and they will be a compilation of each group member's responses. P2-4 will conduct an assessment that involves both the students asking questions about the listening text and the teacher asking questions. That is to say, she will conduct an integrated assessment consisting of listening and speaking skills.

In summary, these findings above revealed that the junior PSETs had indicators of weak and strong literacy regarding the assessment of young language learners' listening skills. However, it is possible to claim that some of them (P1-2 and P1-3) have more indicators of

weak literacy. In comparison to them, the others (P1-1 and P1-4) are more literate about assessing the learners in this group.

Assessing the Listening Skills of Adolescent Language Learners

The group of adolescents refers to the students studying at an upper secondary school. Similar to the ones for the group of young language learners, the members of the first case also have some indicators of *weak literacy* regarding the assessment of the adolescents' listening skills. These indicators consist of item types such as multiple-choice questions, fill-in-the-blanks, matching, and true-false statements.

Table 10

Assessing the Listening Skills of Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
- Multiple-choice questions (P1-2, P1-3, P2-1, P2-2) - Fill-in-the-blank statements (P1-2, P1-3, P2-1, P2-3) - True-false (P2-1, P2-2, P2-4) - Assessing with a written exam (listening to a native speaker) (P1-2, P2-5) - Matching (P2-1, P2-4) - Question and answer (Paper-based) (P2-4)	- Asking questions about the short film, video, song, or series (P1-1, P1-3, P1-4, P2-1, P2-2, P2-3, P2-4) - Role-play (P1-1, P1-4, P2-3, P2-5) - Assessing listening and speaking (Listen to a text and talk about the topic) (P1-3) - In-class discussions after listening to recordings about various topics (P1-4) - Group assignments (Writing or speaking after listening) (P1-4) - Listening and Speaking (Answering the teacher's question) (P2-1) - Listening and Speaking (Pair work or Group work activity, Commenting on a listening track) (P2-1) - Listening and writing (Completing the listening track, Commenting on the listening track, Answering a question in the track) (P2-1) - Asking students to detect objects after listening to instructions (P2-1) - Making interpretations from the comprehended parts (P2-2) - Students drawing lines on the map between places visited by the group (P2-2) - Students circling the places the group visited on a map (P2-2) - Students writing the type of activity around the circles on the map (P2-2) - Asking students to retell what they listened to in short and simple sentences (P2-3)

-
- Assessing listening and writing (Writing a text/Filling out a table according to audio) (P2-3)
 - Assessing listening and speaking (Students asking questions about audio/Describing some locations) (P2-3)
-

P1-2 mentioned the items that are appropriate to assess the listening skills discretely, although he commented that the language skills should be assessed in an integrated way since they are interconnected. Besides, he is concerned about adopting assessment methods that involve constructed responses. Regarding this, he states that ensuring objectivity in the scoring of the outputs produced through these methods is not much of a possibility. Therefore, he prefers to use selected response items more. Additionally, he writes that he will administer a pen and paper exam in which his students will listen to an audio recording of a speech made by a native speaker. Likewise, P1-3 mentioned these tools of assessment in her answers to the questions asked in the semi-structured interview and the scenarios.

I plan to do it with a more objective question style and more objective question types like multiple-choice questions and fill-in-the-blanks (P1-2, Semi-structured Interview III).

By having students listen to a recorded video, we can ask them open-ended questions, and we can also evaluate them by filling in the blanks (P1-3, Scenarios for Foreign Language Assessment).

The student teachers also mentioned the practices that are more compatible with integrated and communicative language assessment or appropriate for the authentic assessment of listening skills. Acknowledged as indicators of *strong literacy*, these practices involved either the discrete assessment of listening skills with real-life tasks or the assessment integrating listening and speaking or writing skills.

Asking comprehension checking questions about the content of the audio or video is one of the most common techniques that the participants in this case will utilize. After having their students listen to and watch songs, podcasts, and videos, the teacher trainees (P1-1, P1-3, P1-4, P2-1, P2-2, P2-3, and P2-4) will ask these questions verbally or in a written way. Stating that he will prefer to use the listening materials on more scientific topics since the adolescents will have more knowledge about grammar and vocabulary, P1-1 mentioned the integration between the listening and speaking skills.

They already start a little in high school. Let's not say academic English, but their grammar knowledge goes much further. As they have seen for about 6,7, and 8 years, their vocabulary is expanding a lot. Because we have units on almost every subject. I would like to have them listen

to some more scientific stuff. For example, there are short films without subtitles. I would open a scientific short film and also ask questions about it to test how much they could understand. I would open a scientific short film and also ask questions about it to test how much they could understand. I would show them a scientific short film and also ask questions about it to test how much they could understand. That's how it will happen. In my opinion, speaking also goes together here, and we need to combine it here. To find out what the child understands in English. Does he understand those words correctly? Can he use those words correctly? I would like to learn about them (Semi-structured Interview III).

As another method of integrated assessment, P1-1 wrote that he would create a role-play scenario in which the students would ask and get answers about the famous places of the country that they visited. Putting emphasis on the fact that the situation given in the scenario is related to real-life situations, he would focus on whether the student could understand the given details and ask for follow-up questions.

Like P1-1, P1-4 also emphasized that she would adopt the listening tracks on more 'sophisticated' topics such as an event in history, artificial intelligence, or issues related to the field of ELT. Claiming that similar themes such as holidays and school years are covered in the course books every year, she wants to go beyond them because the students are more interested in the other topics. After the students listen to the audio, they will discuss the topic in the classroom with their peers in light of the teacher's questions and write a text based on the listening with their peers as a group assignment. As the reason for using this assessment practice, she states that the multiple-choice questions and fill-in-the-blanks involve the chance factor, and this hampers the assessor's ability to diagnose whether the students could understand what they listened to or not. Furthermore, it is important to note that in her answer to the question in the scenarios, P1-4 wrote that she would assess her students' listening and speaking skills together with role-plays and discussions.

In contrast to these ways of assessing language skills, P1-4 said both in the third semi-structured interview and in her answer to the scenarios that she liked to assess the learners' language skills discretely. For these reasons, she stated that a discrete assessment would be more ethical and more appropriate for getting detailed results about each skill. So, this indicates that she has confusion about the ethical assessment process and the appropriate assessment practices.

Drawing on her experiences, P1-4 criticized the fact that listening skills were not given enough attention in the primary and secondary schools and were taking up only 10 points to complete 100 points in the exams. Therefore, she wrote that "I can clearly state that students

have a great deficiency in terms of listening skills.” Also, she noted that this incompetency should be revealed with new assessment contexts and tools with a process-oriented view.

Instead of being evaluated by listening to a dialogue and answering 5 questions about it, I would like to progress by learning what the song tells, by writing the words or phrases that come to mind after listening to a song, or by learning if the pronunciation of the word in this song is different from what is believed or what other meaning it has than what is known, and I would like to make an assessment by looking at all of these processes. Moreover, I would encourage students to be involved in these vocal activities, such as singing, theater performances, etc., and I would include them in the assessment (Opinion Journal IV).

In summary, the junior pre-service English teachers are planning to employ not only the traditional tools of assessment, which are less related to real-life activities and less conducive to language use, but also the communicative assessment tasks, which are more authentic and appropriate to assess integrated language skills. Therefore, it should be noted that although the student teachers are juniors in the English language teaching program, they generally have the knowledge that the listening skills of the adolescent students should be assessed communicatively.

Assessing Speaking

Assessing the Speaking Skills of Young Language Learners

The junior PSETs’ literacy in assessing the speaking skills of YLs is high. They plan to deploy assessment methods conducive to both speaking individually and interacting with their teacher or peers.

Table 11

Assessing the Speaking Skills of Young Language Learners

STRONG LITERACY
<i>Interaction</i>
• Dialogue between peers (P1-1, P1-2, P2-1, P2-4, P2-5) • Role-play (P2-1, P2-2, P2-3, P2-4) • Question and answer (P1-4, P2-1) • Discussion (P2-5)
<i>Production</i>
• Making a presentation (P1-2, P1-3, P2-2, P2-4) • Introducing their families in a short speech (P1-2, P2-2, P2-3) • Giving directions or describing the distance from a place (P1-1) • Describing the place of an object to a peer (pairwork) (P1-1) • Talking about their age, siblings, etc. (P1-1)

-
- Describing a given picture (P1-3)
 - Giving a topic to the student and asking them to say a sentence about it (P2-1)
 - Reading and speaking assessment (P2-1)
 - Describing their parents and comparing with peers' (P2-2)
 - Asking the students to talk about their hobbies, likes, and the places they want to travel to (P2-3)
 - Listening and answering the questions (P2-4)
 - Talking about what s/he did on vacation (P2-5)
 - Expressing one's own opinions about a topic (P2-5)
 - Telling a tale by La Fontaine after reading the book (P2-5)
-

The junior PSETs reported that they would assess their future students' speaking skills through various assessment practices. Requiring the students either to interact with their peers or teacher or to make a speech by using English, these practices were grouped into two main terms: interaction and production. The practices under the term interaction involve a dialogue between peers or between the teacher and the student. On the other hand, production refers to the speech that the student makes by herself or himself about a topic with or without preparation. A common feature of the practices given under these terms is that they require the learners to use the target language in a communicative way and thus allow the teacher, as an assessor, to make direct observations on the students' speaking competence. In this section, the assessment practices having this feature are called indicators of a *strong* level of literacy in assessing the young language learners' speaking skills.

As in assessing listening skills, P1-1 drew on his experiences while answering how he will assess his students' speaking skills. Moreover, he took into consideration what the young language learners were able to do with English and mentioned an assessment practice allowing interaction between the students.

In speaking, at first, at the beginning, we again take the second and fourth grade range. Here, we have given the things that children understand, for example, the rules, or we have already taught them whether they listen or not with songs like this. How much can he apply these while talking to his friends, for example? I am not talking about a second narration, but I am thinking about how to explain it, with your permission. For example, here are the things they learned: They could tell how old they were. They could tell how many siblings they had. They were able to give simple directions. Turning them into a group study activity. For example, there are 10 students in my class. For example, we had something. When we were in the fourth grade, an eraser was placed in a random place. He was asked to describe it. I would like to do something like that and see that they convey what they learned in a comfortable way without paying too much attention to the student's pronunciation. In other words, I correct very big mistakes, and by not correcting every word, I would look at how well the student could explain without breaking the student's confidence or breaking his happiness (Semi-structured Interview III).

For the lower secondary school students, P1-1 expressed that he would expect the learners to make sentences with a little more detail. However, he also added that it was not very possible for them to speak fast and fluently because they did not have enough lexical and grammatical knowledge. As a task, after the teacher gave a sketch to the students, they would give directions for describing the place of a building or the distance of a place to somewhere.

Similarly, P1-2, P2-1, P2-4, and P2-5 stated that they would like to assess their students' speaking skills by asking them to interact with their peers. Of these participants, P1-2 expressed that he was always in favor of the spontaneous assessment of speaking skills regardless of the learners' level because he thought that such an assessment was of better quality. Moreover, he noted that it is important that the students have enough knowledge about the topic, whether the assessment is integrated or not. With this in mind, P1-2 pointed out that he would ask his students to have a conversation with each other by asking and saying their name and surname, where they are from, and how they are.

In the Opinion Journal 4, P2-1 also wrote that he would assess his students' speaking skills through dialogue with their peers. In these dialogues, he expects the learners to communicate with simple sentences and later to ask and answer more complex questions as they progress. However, to use this technique, P2-1 thinks that the students should have basic knowledge of grammar and vocabulary. For this reason, he is planning to first teach grammar and vocabulary to his students and then assess their knowledge through traditional testing tools such as matching, fill-in-the-blanks, and true-false.

To assess the students while they are interacting with each other, P2-4 wants to organize speaking clubs during certain class hours or on certain days of the week. During these meetings, the students will talk about a topic that they will jointly decide. Moreover, she draws attention to the positive impacts of these meetings by stating that the students will socialize and be encouraged to speak English.

As compatible with what P1-1 said, P2-5 expressed that the student's not being demotivated while speaking was an important issue for him, and therefore he would pay more attention to speech intelligibility. Also, he added that the interaction between the students would reduce their anxiety and motivate them to speak because they would have time to think about what they would say while their peer was speaking. Stating that the assessments would generally focus on the interactions between the students or the student and the teacher, P2-5

would like to assess his students by asking them to speak among themselves about a topic covered in the lesson.

Role-play is also another prominent assessment technique that the junior pre-service English teachers plan to use for assessing the young language learners' speaking skills. Here, it should be noted that only the participants from the second grade mentioned role-playing as an assessment technique. However, there is not any data to explain why the other junior teacher candidates did not want to employ it.

As for how they would draw on the role-play to assess speaking, the participants have different opinions. For example, P2-1 wrote that he would ask his students to write a script with their friends in small groups. The script will include the taught expressions of asking and giving permission and requesting. Then the students would act out their roles by adhering to this script. If they have difficulty remembering the text, they can improvise, but looking at the text during the drama is forbidden. Moreover, speaking unrelated to the topic will cause the student to lose points even if s/he makes a very fluent speech.

With the aim of having his students speak spontaneously, P2-2 stated that he would create a task where they would randomly choose a character, a modal verb, and a verb. Emphasizing that fluency and purposeful communication would be the most important two criteria, as well as vocabulary, grammar, and pronunciation, P1-2 wrote the following explanation about how he would conduct an assessment on speaking.

To keep it as original as it can be, I would prepare a list of roles (parent, friend, teacher), questions (can, may, could) and answers, and verbs (open, close, go, leave, etc.) on paper and cut each item into individual pieces. Then, I would group them in three different bowls and have students come up and pick an item from each bowl. According to the items they pick, they can role-play, create original utterances, and remain spontaneous while allowing the language to be used for novelty. The items would consist of the words that have been taught in the previous lessons (Scenarios for Foreign Language Assessment).

In the scenarios, P2-3 also gave the answer that she would use the role-play as an assessment technique. She wrote that she would divide the class into pairs or groups of three by taking class size into account. Then she would give each pair or group a task that involves situations that the students can experience in daily life and requires them to use short and simple permission and request statements for these situations. Moreover, P2-3 added that if the students made errors that hindered communication and thus prevented the assessment from achieving its aim, she would intervene in a way that did not disturb or make them anxious in order for the students to correct their errors.

The junior teacher trainees stated they would also apply some assessment practices that require the young learners to give a speech. That is to say, with these practices, the learners are aiming to share their opinions or give information about a topic. Making a presentation is the most preferred of these. Regarding this, P1-2 briefly pointed out that he would ask the lower secondary students to make a presentation about various topics. He added that the students can make a bit more complex sentences because they are more familiar with English.

In his Opinion Journal IV, P2-2 wrote about speaking, saying that “this skill may be one of the most flexible yet also stressful skills to test between all four”. Emphasizing that “active, spontaneous production” may be challenging for the students, P2-2 wrote that “the students can feel stressed out, ashamed of their proficiency, or pressured”. Therefore, he commented that “especially for younger learners, speaking skills can be based on pre-planned speeches about simple topics (e.g., routines, hobbies, introduction of family tree)”. In line with this view, he wrote this explanation in the scenarios for assessing both primary and lower secondary school students.

If such an assessment is planned for a later date, I can ask my students to prepare a family tree as a poster with their parents’ pictures and a few adjectives written beneath the pictures. The students can hang their posters on the board and talk about their family members. The sentences do not have to be complex; using the correct subject and adjective in the sentence would be more than enough (Scenarios for Foreign Language Assessment).

As for the presentations that will be made by the lower secondary school learners, they will be given in two ways. First, the presentations will be given by the students to introduce themselves, and they will last 5 minutes. Second, after listening to the audio, the students in groups will prepare a presentation consisting of the answers asked about the content given in the audio.

Drawing on her experiences, P2-4 would like to ask her prospective students to make presentations about the topics they choose. She thinks that this task will encourage the students to speak English better, as expressed in the sentences below.

Instead of subjecting my students to evaluation criteria in the form of written exams, multiple-choice questions, and true-false, I would ask them to make presentations within the framework of certain topics they wanted, just like my teacher did to me. In this way, I think that the student will want to explain something that s/he is interested in and will be able to speak better at this point (Semi-structured Interview III).

In addition to giving a presentation, the junior teacher candidates would have their students speak by either giving them various prompts or topics or asking them to prepare a poster or

a table. Related to this, P1-3 pointed out that she would give a picture to the young learners and ask them to describe it. P2-3 remembered that she was expected to speak to give personal information when she was a young learner. As compatible with this, she stated that she would assess the primary school students by giving them a theme like hobbies, what they like, and the countries they want to visit. Likewise, P2-5 expressed that he would ask his students to talk about what they did on vacation. For the lower secondary students, he prefers to conduct an integrated assessment in which the students read a tale by La Fontaine and tell it in very simple sentences. Stating that he would ask his students to draw pictures of their parents, P2-2 explained in the scenarios how he would conduct an informal assessment in the classroom.

If the activity is planned to take place in the same lesson, students can instead draw their parents on a piece of paper, exchange the papers with their classmates, and make comparisons. Then I can ask them to first describe their own parents and compare them with the parents of their classmate who is sitting next to them (Scenarios for Foreign Language Assessment).

Assessing the Speaking Skills of Adolescent Language Learners

In this research, the adolescent language learners consist of the students studying at the upper secondary school. For assessing the speaking skills of the students in this group, the junior pre-service English teachers would like to adopt communicative and integrated assessment methods. Making a presentation, role-playing, having a discussion, and asking the student to talk about a topic are the frequent methods that will be used.

Table 12

Assessing the Speaking Skills of Adolescent Language Learners

STRONG LITERACY
• Making a presentation (P1-2, P1-3, P2-1, P2-2, P2-3, P2-4) • Role-play (P1-2, P1-4, P2-1, P2-2, P2-3, P2-4) • Asking the student to talk about a topic (P1-1, P1-2, P1-4, P2-3, P2-5) • Integrated assessment (P1-3, P2-1, P2-3, P2-4, P2-5) • Discussion (P1-3, P1-4, P2-3, P2-5) • Speaking in pairs (P2-1, P2-5) • Speaking in groups (P2-1, P2-4) • Introducing the places in a plan to the other group (P1-1) • Giving a topic to the students and asking the students to talk with their peers (P2-1)

Making a presentation is the method that both the first and second graders will prefer to use. All of them mentioned this assessment practice during the third semi-structured interview. Of these participants, P1-2, P1-3, P2-1, and P2-4 stated that they would like their students to give a short and simple presentation about a topic they choose. Expecting that the students can give a bit longer presentations, P2-2 also expressed that “at high school age, although it is no longer very complex, it is not oriented towards a specific field like ours, but a presentation can be made on a certain theme, again based on the theme in the textbook”. As different from these, P2-3 would ask her students to do a search on a topic outside the classroom and make a presentation about it in the classroom.

Role-playing is another assessment practice most preferred by junior pre-service teachers. It involves the learners acting out according to a script written by them or a prompt like a story or a video. P1-2 wrote in the scenarios that they would assess “the speaking skills by acting out a story”. Preferring his students to perform a story, similarly, P2-1 noted how he would carry out the assessment as given below.

I would divide the students into groups of 3 according to their levels and give them a story to act out. Then I would give each group 3 days to get prepared (since it is not possible for every group to do it on the same day, I would give the remaining groups their stories later so that they have equal preparation time). Each group would have 20 minutes, of which the first 10 minutes would be for acting out the story, and the rest of the time I would give them time to talk about a tradition of their choice. The evaluations would be recorded, and the students would be informed in advance (Scenarios for Foreign Language Assessment).

As for P2-2, he wants to conduct an assessment involving presentation and role-play. Aiming for the students to randomly select the roles they would perform, P2-2 explained the assessment he imagined to be done on speaking skills as follows:

First of all, I would request that the learners form groups of three or four. Depending on the number of groups, students can pick cards that represent different traditions (like a raffle). Depending on the theme they receive, students are asked to create a role-play that combines acting with presentation skills. Within a specific time period (minimum 10, maximum 15 minutes), students can present information for a minimum of five minutes and then proceed to act. It is essential for all group members to have at least one common role, so each student should have a part in the presentation even if they are going to act it out later (Scenarios for Foreign Language Assessment).

Emphasizing that she will create a friendly and comfortable environment for her students, P2-3 also wants to employ role-play as an assessment method. In addition to how the learners use the target language, she will also take such factors as body language, tone of voice, and facial expression into consideration. However, it should be noted that this involves the risk of constructing irrelevant variance, and this practice is considered a sign of weak literacy in scoring.

In the Opinion Journal IV, P2-4 wrote that she would ask her students to create dialogues by using the structures they were taught and acting them out. Besides, she noted that she would “show her students clips from movies, TV series, and animations and ask them to act out the scenes and record themselves on video or audio.”

The junior teacher trainees also preferred discussion. This assessment practice means that the students talk about a topic and exchange ideas in groups. The participants who would like to use this alternative assessment method provided various explanations. For example, in her Opinion Journal IV, P1-3 stated that her students would discuss topics related to daily life. Also, she added that the discussion would be recorded in case the students objected to their scores. Drawing attention to the importance of context, P1-4 had the idea that the students should be provided with a context that would allow them to speak as given below.

At the high school level, I would make it possible to speak after activities with a clear purpose, because suddenly speaking can become a challenge for students. For this reason, what I just mentioned can be a dialog, a text, or a storytelling podcast. Since the variety has already increased a lot, it is not very difficult to find such content today. I think there would be a discussion on that. I think that's how I would evaluate it (Semi-structured Interview III).

P2-3 wrote that “I plan to use alternative methods to assess all my students' English-related skills”. In addition, she believes that “instead of uniform and standardized exams, assessments that are spread over a wider period of time and intertwined with life will be better for [her] students' language development”. Similar to P1-3, she wants to organize discussions on topics that are directly related to real-life, and for this, she will “divide [her] students into groups, give them specific topics, and ask them to create a decent discussion on these topics”.

Asking students to express their opinions about a topic they choose or are given by the teacher was also mentioned by the members of the first case. In this regard, P1-1 remembered that they talked about the topic of “love” when he was a high school student. Stating that he will allow the students to talk about the topic they want, P1-1 pointed out that this speech will be 2 or 3 minutes long.

P1-4 wrote that she would encourage her students to speak in English rather than their mother tongue. Besides, she noted the idea that process-oriented assessment of speaking skills would be fairer and more awareness-raising and explained how she would assess her adolescent students' speaking skills with these words:

At the end of each lesson, I prefer an assessment method that covers a whole semester instead of an assessment done on a single day. Observing the students with activities such as five people selected that day talking about a short story or event (this can be on any subject, as long as the students can express themselves comfortably), reflections about the TV series or movie watched, comments on the book read, etc. Because I believe that this will both raise more awareness and produce a fairer result (Opinion Journal IV).

P2-3 plans to carry out an informal assessment integrating reading and speaking skills. In this respect, she wrote that she would bring realia such as newspapers and magazines to the classroom, and then she would ask her students questions about the content presented in these materials. Thus, she will “have the opportunity to assess [her] students’ speaking skills in relation to daily life”.

Integration of listening and speaking skills was mentioned by P1-3 in her answers to the scenarios. After the students listened to the audio, they would express their opinions about the content of the listening text. Likewise, P2-1 also had the idea of integrating the listening, reading, and speaking skills. With regard to this, he wrote that her adolescent students would complete a half-given listening audio or reading text.

P2-3 would have her students ask questions about the listening text or describe the location of a place by drawing on the audio. P2-4 also wrote in the scenarios that “by integrating reading with speaking, I would ask my students to do a short speaking activity about the text they read”. Similarly, P2-5 stated he would carry out an informal assessment by having them discuss a topic given in the text brought by the teacher to the classroom. Additionally, he expressed that this would allow the students to interpret the text from different perspectives and criticize it.

In the scenarios, P1-1 wrote that he would divide the students into groups and distribute sketches to each group. Then the students in a group would ask the other group about the location of a place, and they would describe the place according to the sketch. In a similar vein, P2-1 emphasized that he would rather allow his students to interact with each other than to answer the questions he asked. As given below, the reason for this is that he thinks the students at the upper secondary school level have enough knowledge about the topic provided by the teacher, and therefore there is no need to ask questions for guidance.

By the time they reach high school, it will no longer be a question-and-answer pattern. Because now, when I ask students to talk about this topic, they probably won't expect a direct question from me. I don't add them there because I don't know if they will be there at A2 level. At B1, B2, and high school level, let me say that instead of asking direct questions to the students, I give them a topic, a subject to talk about, not with me but with their classmates. In A1, A2, of course, these can also be used, like when you ask your friend a question and he answers you, but the

teacher can do this because only one side asks questions. At the high school level, group conversations can be used as an assessment. Because now I think they can speak fluently and on a specific topic for at least a few minutes—not a few minutes, but as long as they want if they are active enough (Semi-structured Interview III).

As in the lower secondary school, P2-4 pointed out that she would also organize sessions of speaking clubs for assessing the adolescent students. In these sessions, they would talk about the topics they chose.

Assessing the Reading Skills

Reading skills involve the students' being able to detect the main idea and the details of a text. The findings are again presented for two groups of learners: young learners and adolescents. However, since the assessment of the young learners' reading skills is not required in the curriculum until the fifth grade (MoNE, 2018, pp. 1-96), the findings given in the following section should be considered for the lower secondary school students in the fifth grade and onwards. On the other hand, the improvement of reading skills is expected for all the students in the upper secondary school, regardless of their grades. Therefore, there is no need for any caution like this.

Assessing the Reading Skills of Young Language Learners

The junior pre-service English teachers mentioned the assessment practices that are indicators of weak or strong literacy in assessing the reading skills of lower secondary school learners. The weak literacy indicators consist of non-communicative or traditional assessment practices that do not much require the learners to comprehend a text in line with a purpose similar to the one in real life. In contrast, the strong literacy indicators consist of communicative, integrated, or alternative assessment practices that require the learners to comprehend the reading text as compatible with communicative purposes and real-life situations.

Table 13

Assessing the Reading Skills of Young Language Learners

WEAK LITERACY	STRONG LITERACY
• True-false statements (P1-2, P2-1, P2-3, P2-4) • Multiple-choice questions (P1-2, P2-1, P2-3, P2-4) • Fill-in-the-blanks (P2-1, P2-3, P2-4) • Asking questions about the main idea, feelings, and events (P1-1) • Asking questions with definite answers (P1-1) • Matching (P2-1) • Yes / No questions (P2-2)	• Open-ended questions (P1-3, P2-2, P2-3) • Integrating reading and writing (P1-1, P1-3) • Discussing the text (P2-2, P2-5)

Multiple-choice questions, fill-in-the blanks, matching, yes/no questions, and questions with definite answers are the constituents of weak literacy. The multiple-choice questions were the tool that three participants wanted to employ for assessing their reading skills. Regarding this, P1-2 stated that “I plan to ask multiple-choice or true-false questions when I start practicing my profession” and “I will try to stay away from using questions that are not classical multiple-choice”. This is because he thinks that they are “of better quality,” and he also believes the constructed-response items have the risk of giving unclear and unethical results because they do not have a clear response.

P2-1 was another teacher candidate who would like to use multiple-choice items. For this assessment, he plans to give a 90- or 100-word paragraph that he will write or that is given in the coursebook and consists of unconnected sentences. He does not want to ask constructed response items to his students because “[he] does not think that they can answer an open-ended question in a long way, in a desired way”. Rather than this, since the students can answer the questions with a clear answer at most, he prefers to use the multiple-choice items which, have 3 or 4 options. In addition to these items, he would like to adopt true-false statements, fill in the blanks, and matching. P2-1 pointed out that he would not ask his students to write anything after reading the text if scores were going to be assigned. Related to this, he stated these sentences below:

Because if I am assessing the reading and writing skills at that level and I use open-ended, I am also assessing writing to some extent. But that wouldn't be the right criterion for me. I would put it there to see their progress, but I think reading skills would still be assessed through reading. Because the maximum answer a child can give is one or two sentences to a clear question, in A2, a child may not be able to write a 50-word answer such as What is your comment on this? and What is the answer to this? So, the reading passage plus these 4 tools would be in use again in middle school (Semi-structured Interview III).

Writing that “I can include materials related to daily life in the classroom environment”, P2-3 expressed that “I can prepare questions about the content I have identified in these

materials”. However, these questions will consist of multiple-choice, fill-in-the-blanks, and true-false items. Although some of the assessment practices she mentioned did not always align with this idea, she expressed that the discrete assessment of the language skills would give more accurate results, as given below.

When assessing one skill, the student may not be able to fully express what is really wanted when the other skill is involved. In order to avoid this, it seems more reasonable to me that when assessing only one skill, it should be done mostly or, if possible, completely in a way that measures that skill. Otherwise, the student may not express himself or herself sufficiently. The assessment may also not be accurate (Semi-structured Interview III).

As for P2-2, he stated, “I would ensure that reading was understood. Interpretation may be too early. It will not take place in lower secondary school classes. Let them understand the text they read”. As these remarks imply, he does not prefer to employ assessment practices that allow the learners to reflect on the text. In line with this view, P2-2 wrote in Opinion Journal IV that “different types of questions can be inserted below the text”. However, these would be in the structured response format, such as yes-no questions, which would require the students to read the text and answer “yes” or “no” without using higher-order thinking skills. In addition, questions about the specific details in the text would also be asked of the learners, and these would be “written in a paraphrased format of the original sentence” existing in the text.

The junior pre-service English teachers also mentioned the assessment practices deemed to be strong literacy indicators. For example, three participants would use open-ended questions that allow the students to analyze the text and express their reflections. These participants, P1-3, P2-2, and P2-3, would like to ask these questions about the reading text, which would not be difficult or too long. Moreover, they added that the topic of the text should be appropriate to the students’ level. Regarding this, P2-2 also stated that the texts should be more about concrete objects like animals, cities, and buildings.

Moreover, some participants want to integrate reading with another language skill. For example, P1-3 stated he would integrate reading and writing. After giving the students the reading text, they would write a short paragraph summarizing what they understood from the text. Similarly, P1-4 wants to conduct an integrated assessment. This would be done through assignments in which the students would write their short reflections about the text.

P2-2 and P2-5 would also like to integrate reading with speaking skills and ask their students to discuss the text they read. P2-5 expressed that through this assessment practice, he would

be able to detect “how much the student understood what s/he read” and “how much s/he presented in a different way what s/he understood”.

Assessing the Reading Skills of Adolescent Language Learners

The junior pre-service teachers referred to the practices that were classified under the categories of weak or strong literacy indicators for assessing the adolescents, as was the case for young language learners. The category of weak literacy indicators involves the traditional assessment techniques: true-false statements, multiple-choice questions, fill in the blanks, and matching. In contrast to this, the category of strong literacy indicators involves the integration of reading with speaking or writing, ordering the events mentioned in the text, and asking questions about the gist and details in the text.

Table 14

Assessing the Reading Skills of Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• True-false statements (P1-2, P2-1, P2-3) • Multiple-choice questions (P1-2, P2-2, P2-3) • Fill-in-the-blanks (P2-1, P2-3) • Matching (P2-1) • True-false statements (P2-2) • Not using much open-ended questions (P2-2) • Reading aloud (P2-3)	<i>Reading and Writing Integration</i> • Writing a summary about the text (P1-3, P2-1) • Open-ended questions (Reading and writing - Integrated) (P2-1, P2-2) • Asking the students to write a paragraph about inferences from the text (P1-3) • Reading and answering open-ended questions with long answers (P2-1) • Asking students to read the texts quickly and take notes (P2-1) • Asking students to write an analysis to detect important points of the text (P2-3) <i>Reading and Speaking Integration</i> • Open-ended questions (P1-3, P2-1, P2-2, P2-3) • Talking about the text analyzed and the order of events (P1-1, P2-4) • Preferring Wh-questions rather than multiple-choice (P1-1) • Asking questions about the specific information in the text (P1-1) • Paying attention to the student's noticing important points and making interpretations (P1-4) • Discussion (P2-2) • Making inferences, criticisms, and comments about the text (P2-4) • Making a short speech about the reading text (P2-4) • Interpreting the text from different perspectives (P2-5) • Criticizing the text (P2-5)

True-false statements, multiple-choice questions, and fill-in-the-blanks were the constituents of the indicators of poor literacy. Stating that he would use more complex texts in comparison to the ones used for the lower secondary school, P1-2 preferred these items rather than the open-ended items.

P2-1 also wrote the structured-response items in the scenarios for assessing the adolescents' reading skills, as in the following sentences:

For reading, students will be given an A2-level text involving general information about their city. An assessment sheet consisting of fill-in-the-blanks, matching, true-false, and short open-ended questions will be prepared for these texts. The questions will include new vocabulary that the students have learned so far, and these words will need to be known in order for the questions to be answered. Students will be given one class hour (40 minutes) to complete the assessment (Scenarios for Foreign Language Assessment).

Although she wrote in the Opinion Journal that she would employ the alternative assessment methods for assessing all language skills, P2-3 expressed in the semi-structured interview that she would use true-false statements, multiple-choice questions, and fill-in-the blanks in the written exams. Besides, she noted in the scenarios that "I can assess their reading skills by having them read a text I have chosen for reading assessment in class or by asking them to make audio recordings on their own". That is to say, P2-3 thinks reading aloud is an appropriate technique for reading assessment. However, this can be a source of construct-irrelevant variance, which occurs "when the test measures characteristics, content, or skills that are unrelated to the test construct" (Reynolds et al., 2021, p. 189), since reading aloud does not give evidence of whether the learner can comprehend the text or not.

The assessment practices constituting the strong literacy indicators, which enable the students to do practices by comprehending the text or to produce writing or speech by drawing on the reading text, would also be used by the junior teacher trainees. The integration of reading and writing skills is one of the most commonly cited practices. In this regard, P1-3 said in the semi-structured interview that she could ask her students to read the text and then write a paragraph summarizing what they understood. In the scenarios, she also re-emphasized this idea and wrote the sentences below:

I think they can be integrated. For example, in reading skills, let's have a topic about a neighborhood or a city. We can assess it with open-ended questions. In writing skills, they can be asked to write a summary or paragraph about the text they read (Scenarios for Foreign Language Assessment).

In a similar vein, P2-3 also thinks she should have her students write an analysis of the text she will give about past habits and experiences. Although she expressed that a discrete

assessment of the language skills would be better for getting more accurate results and improving them more, she preferred an integrated assessment for assessing reading skills in the scenarios.

P1-4 emphasized both in her Opinion Journal and during the interview that she would give reading assignments to her students and ask them to answer the questions “to put forward [their] own opinions based on the text”. Preferring process-oriented assessment rather than “an exam involving all kinds of skills”, P1-4 wants to assess the learners’ reading skills through these assignments that will be given “once or twice during the semester”.

Can it be six or seven texts in the form of homework? It can be long. These would be large-scale assignments anyway. I ask questions to measure who knows what, and it's usually not that kind of answering that you pick and choose with tweezers. I would prepare a reading assignment that requires more brain power, where there are points that really need to be criticized. Yes, let them explain what the text says. I wanted them to strengthen their vocabulary and realize something, but the questions I would ask would not be simple. I mean, I wish there were things like this where the student could really add his or her own interpretation, and at the same time, I would like to see a contribution in writing skills while responding to that proposal (Semi-structured Interview III).

P2-1 also integrated reading with writing and expressed this in the interview: “In high school, if possible, I would try to remove these four I mentioned: matching, true-false, fill-in-the-blanks, and multiple-choice questions if possible”. However, he also thinks that the level of the students should be taken into consideration when doing this. And he adds that “I would remove these and instead try to introduce more difficult open-ended questions, reading and writing intensively”. As a response to these open-ended questions, he would “expect more accurate long answers from the children in the writing part as [he] would be assessing more writing”. In addition, he wrote in the scenarios how he would assess reading and writing together.

I would give the whole class biographies of two different famous people and texts to study. First, I would ask them to read the text quickly in a limited amount of time and then note down the main points (skimming and scanning). Then I would ask them to summarize these texts on a blank sheet of paper in a limited amount of time. If the students are competent enough, I would also add a word limit. In the section on celebrities, I would ask the students to read the biographies quickly, take notes on the main points, and summarize the lives of these people, including important moments in their lives. To summarize, I would prepare an assessment that includes the ability to use reading techniques, the ability to take notes, and the ability to summarize. The abilities mentioned here are reading techniques, note-taking, and summarizing, respectively, as they exist in the literature (Scenarios for Foreign Language Assessment).

P2-2 wrote that he would ask opinion-based questions of his students in a written exam. The reading text that would be given in this exam would be 150-175 words long. He explained these questions in the scenarios:

The opinion-based questions would require students to share their own ideas in the target language. Questions such as “Do you think <...> was a good habit?” “How do you think <...> experience contributed to the person?” would be suitable examples for this category. Although there may be no “definite” correct answer, as long as students’ answers are thematically and structurally correct, the question remains a rewarding addition rather than a punishing one (Scenarios for Foreign Language Assessment).

Reading and speaking integration is another sub-dimension of strong literacy in the assessment of reading skills. It involves the students talking individually or with someone based on the content of the text. In this context, P1-1 wrote that his students would talk about the work they analyzed and the biographical sequence of the given events. On the other hand, P2-2 and P2-5 would organize discussions and dialogues among his students as informal assessments. Moreover, according to P2-2, the speaking sections would occur after they read the texts, which would be about not only concrete elements but also abstract topics.

The teacher candidates also mentioned integrating reading and speaking skills. While doing this, P2-4 remembered the reading assessments she experienced. Instead of multiple-choice questions, which she called ‘test’ and true-false statements, she wants to integrate the two skills in the informal assessments she will carry out.

We were asked to read something. We read it, and then we were asked certain questions about what we had read in order to measure whether we understood it or not. I probably would have gone for that kind of assessment, but as I said, it wouldn't have been just that. I wouldn't measure the student's comprehension with a test or true-false questions. That's where I would make the difference. Instead, I would integrate it with a different skill. It would probably be more speaking. I would want students to use their speaking skills. And I think the lesson would have been a bit like a conversation about their understanding of what they read. This is how I would carry out my assessment (Semi-structured Interview III).

Assessing the Writing Skills

Writing skills involve the learners producing pieces of writing about a topic or as answers to open-ended questions by making grammatical sentences and linking these sentences with appropriate linkers. Here, the following sub-sections involve the findings regarding the young learners and the adolescents.

Assessing the Writing Skills of Young Language Learners

The young learners begin to get training on writing skills in the second semester of the sixth grade. Therefore, the findings presented in this section are related to lower secondary school students. According to the curriculum taught in the 2-8th grades, this group of learners will

be able to prepare posters, invitation cards, and short pieces of writing like paragraphs and stories. In consideration of the objectives in the curriculum, the assessment practices that were mentioned by the teacher candidates were categorized under the categories of weak and strong literacy indicators. The weak literacy indicators consist of assessment practices that are inappropriate to the level of the young language learners. In contrast to this, the strong literacy indicators involve practices that allow the students to produce short texts that are compatible with their English level.

Table 15

Assessing the Writing Skills of Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Paying more attention to grammar (P1-1) • Writing an essay (P1-2) • Explaining factual information with a scientific, historical, or everyday topic (P2-2) • Writing a text about the units or topics the student is interested in (P2-3) • Writing sentences (P2-4)	• Open-ended questions (P2-1, P2-3) • Writing short stories (P2-3, P2-4) • Giving directions according to a sketch (P1-1) • Writing a description of a picture (P1-1) • Writing answers to questions about the names, ages, and jobs of family members (P1-1) • Writing about a given or chosen topic (P1-3) • Giving a writing assignment each week (P1-4) • Searching for a topic and writing about it (P2-2) • Writing about their dreams, wishes, and interests (P2-2) • Writing simple texts about the units (50-70 words) (P2-3) • Writing a story with the words learned and acting it out (P2-4) • Writing a diary (P2-5)

P1-2 stated in the interview that he would assess his students by asking them to write essays. Moreover, he thinks that the third and fourth graders' writing skills can be assessed by the short paragraphs they will write to introduce themselves. P2-2 also plans to assess the learners through assignments that require them to explain factual information in the text they write.

Expressing that she would start assessing writing skills in the 2nd grade, P2-3 mentioned various types of items, such as fill-in-the-blanks and open-ended questions.

I can start with fill-in-the-blank questions because the level of second grade students is lower and they are just starting. I can take it one step further and move on to open-ended questions, but again with short and clear answers—that is, questions that process detailed answers, even if they are not very short. Then again, since there are certain units in 2nd grade, I can ask for simple texts related to those units. Write a text of, say, 50 or 70 words. In lower secondary school, I can increase the word limit by continuing on the basis of units or by having the student write a text on a field of interest. I can ask them to write that kind of text again (Semi-structured Interview III).

In addition to these weak literacy indicators, the junior PSETs would also employ the assessment practices labeled under the category of strong literacy. For example, P1-1 said he would not expect the students to write long texts in the exam and explained how he would assess the writing skills as below.

You wouldn't ask that child to write a long text. When we get to the middle school part, tiny little things start happening. For example, you have them draw a sketch on the exam paper. You ask them to give directions on this sketch, for example. You want them to write 4 or 5 sentences, or when I ask them about their families, I expect them to be able to write me their names, ages, and occupations. I mean, of course you don't expect that child to write an essay (Semi-structured Interview III).

Writing a paragraph about the topic assigned by the teacher or chosen by the learners was a practice that would be used by P1-3. In addition, she also mentioned reading and writing integration and stated that the learners would write their paragraphs based on the reading texts given by the teacher.

In both her Opinion Journal IV and the semi-structured interview III, P1-4 emphasized that she would give writing assignments to her students every week. Instead of the written exams administered within a limited time period, she wants to track through these assignments how her students improve their writing skills. Stating that reading and writing skills could also be integrated, and thus the students would have information about the topic before the writing process, P1-4 wrote:

I would follow a long-term process in terms of writing skills. To be more precise, I believe that this writing skill, whether it is writing a birthday invitation or an essay, should be taken seriously, and I do not support students to follow the examples they see in textbooks because they are so clichéd and rote-based that there is zero gain in terms of academic language. [...] To summarize, at the end of a long-term writing skills training, I would like to see an assignment in such a way that I can actually get a return on the training I have given (Opinion Journal IV).

P2-1 also wants to carry out an integrated assessment in which the students will write answers to the open-ended questions about the reading text. The length of these answers will be approximately 30 words. However, as he notes for assessing reading, this will not be a

formal assessment involving the scoring of the texts written by the learners. Rather than this, it will be adopted as evidence for detecting whether the students are able to understand the text and for providing feedback.

P2-3 said her students would write short pieces about the themes covered in the units or the topics they were interested in. The word limit for these pieces will be more than 70 words. Additionally, she wrote that the learners would write short stories. Likewise, P2-4 will also have her students write this kind of story, the topic of which will be chosen by the learners themselves. In addition to the stories, her students will also write dialogues, and they will act out these dialogues and stories with the classmates they choose. Writing diaries will also be employed by P2-5, as explained in the following statements:

The assessment of writing can be done not through exams, but for example, the student can be asked to keep a diary, and students are often limited in what they do in a diary. So do we, and these limited activities are already at a level that can be explained with short simple verbs or chunks at that level of the student. I think these writing skills of the student can be assessed by keeping a diary or by adding a little bit of their own creativity or point of view to the work, and in such ways, assignment types and formats are important because, when giving such exams to students at this level, especially at a young age, you know, how should I administer them? We force them to get used to the exam a little bit. It actually changes their perspective on language (Semi-structured Interview III).

Assessing the Writing Skills of Adolescent Language Learners

According to the 9th-12th grade curriculum, adolescent language learners will be able to write a series of sentences, e-mails, letters, blogs, stories, essays, and prepare posters, brochures, or leaflets (MoNE, 2018, pp. 1-69). The present research revealed that the junior PSETs would like to employ the assessment practices that were labeled as strong literacy indicators. These practices consist of methods such as writing paragraphs, essays, e-mails, dialogues, or diaries.

Table 16

Assessing the Writing Skills of Adolescent Language Learners

STRONG LITERACY
• Writing an essay (P1-2, P1-3, P2-3) • Writing a paragraph (P1-3, P2-1) • Open-ended questions (Reading and writing- Integrated) (P2-1, P2-2) • Writing a letter of complaint about a scenario that the teacher provided (P2-1, P2-2) • Writing short stories (P2-3, P2-5) • Writing your own ideas about a topic or happiest day (P1-1) • Paying attention: Punctuation, grammar, language use, and vocabulary choice (P1-1) • Assessing reading and writing (Writing about a recent event) (P1-1) • Writing about a topic within the specified time (P1-2) • Writing a paragraph about inferences from the text (P1-3) • Writing a summary or a paragraph (P1-3) • Assessing listening and writing (Open-ended questions) (P1-3) • Assignment (Listening and writing integration) (P1-4) • Writing a 100-150 word text about a given topic (P2-1) • Completing a reading text or listening track (P2-1) • Commenting on the listening track (P2-1) • Commenting on an event in a text given by the teacher (P2-1) • Preparing a script for role-play (P2-1) • Comparing cities in the given table (P2-1) • Writing a speech text (P2-1) • Asking the student to explain their own opinions according to the reading text (P2-2) • Collaborative writing (P2-2) • Writing about and comparing the studies' findings (P2-2) • Asking students to paraphrase the short sentences given by the teacher (P2-2) • Searching for a topic and writing (P2-2) • Writing about some debatable topics (P2-2) • Asking students to prepare a portfolio (P2-3) • Assessing Listening & Writing (Writing a text/Filling out a table according to audio) (P2-3) • Writing petitions or reports (For language majors) (P2-3) • Writing a dialogue and acting it out (Writing, speaking, grammar, vocabulary) (P2-4) • Writing diaries (P2-5) • Writing e-mails (P2-5)

Asking the learners to write an essay is one of the most prominent methods that will be used by the teacher candidates. Related to this, P1-2, P1-3, and P2-3 had similar views, and they expressed that their students would write an essay about a topic given by the teacher or chosen by the students themselves. In addition to this, the adolescent students would write a 100-150 word paragraph about a given topic. For writing such a text, the learners will be given 10 or 20 words that will be used in the writing.

The PSETs would also like to integrate listening or reading with writing skills for assessing the adolescents. For instance, P1-4 would give assignments in which the learners would write a text after listening to audio about some topics such as historical events, artificial intelligence, or language learning. P2-1 also stated that “listening and writing can be done again. Again, continuing the text over the text they listen to or commenting on the text or answering a question in the text by adding their own thoughts”.

In the scenarios, P2-3 wrote “For integrated assessment, I can ask my students to listen to a listening text about cities in different countries or Turkey and write a text or create a table comparing the cities in the listening text”. As given in this quotation, she prefers conducting an assessment integrating listening and writing. In this assessment, the learners will get the necessary information from the audio and make inferences about the cities by comparing them with each other. Thus, the students will use not only their language skills but also their cognitive skills.

Emphasizing the integration of reading with writing, P1-1 wrote in the scenarios “I would give them an article in English about an interesting topic that happened in the recent past and ask them to write their opinion about it”. Similarly, P1-3 wrote in the scenarios and in the Opinion Journal IV that the students could write a summarizing paragraph about the reading text. P2-1 and P2-2 would also ask open-ended questions about the reading text and expect their students to give answers to them by making inferences from the text. Furthermore, P2-1 also stated the learners could write their own paragraphs for commenting on the reading text. He wrote:

I would give three different events that marked the ancient periods to three different parts of the class and ask them to analyze them within a limited time using reading techniques [...] and to interpret them within the limits of their vocabulary by mentioning important points (Scenarios for Foreign Language Assessment).

As another practice, P2-2 wrote in the scenarios that “I would give students multiple short sentences from various topics (sports, music, games, daily topics, and politics) and ask

students to paraphrase them”. Also, he added “For the e-mail or complaint letter part, I would give the details about the context of the complaint and the receiver and ask them to write it within a maximum of 100 words”. Moreover, he pointed out that the learners would write their reflections as given in the sentences below.

Only in the later levels do I assume that the grammatical structure is already in place, that it has found its place, and I start to create activities for it. For example, here, it goes in parallel with reading, but let the child explain what his or her ideas are about a certain topic, or according to the reading above. Here, I realized that they all started to integrate with each other. For example, in the later levels, he can now understand the text he reads; he is still technically bound to the text in terms of content, but he is absolutely free to formulate his answer. He perhaps puts less of his own experience into it this time than at an earlier age. In a fantasy world, a friend's experience is not going to happen anyway. S/he has to stick to the text above. But at later levels, it emerges with the aspect of expressing itself by using its own freedom within the framework of that text (Semi-structured Interview III).

Writing various types of texts was also mentioned by the teacher trainees. Writing a speech text is one of them. P2-1 wrote that his students would write their speech text. Writing short stories was also one of the assessment practices that would be employed by P2-3 and P2-5.

Assessing the Language Components: Young Language Learners

Language components consist of grammar, vocabulary, and pronunciation. However, since the learning outcomes for pronunciation are not given for the primary and lower secondary schools in the curriculum, the findings relevant to the assessment of grammar and vocabulary will be presented in this section.

Implicit assessment of language components is highly recommended in the curriculum (MoNE, 2018). Therefore, the assessment practices enabling the teachers to assess how their students are able to use structures and words are labeled as strong literacy indicators. On the other hand, practices involving the teachers conducting explicit assessment of the language components without requiring the learners to use them are categorized under the category of weak literacy indicators.

Table 17

Assessing Grammar: Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P1-2, P1-3, P1-4, P2-1, P2-2) • Multiple-choice questions (P1-2, P1-3, P1-4, P2-4) • Explicit assessment with a written exam (P1-1, P1-2, P1-4) • True-false statements (P2-4)	• Assessing grammar in writing (P1-1, P1-3, P2-1, P2-2, P2-3, P2-4, P2-5) • Assessing grammar in speaking (P1-2, P1-3, P1-4, P2-2, P2-3, P2-4) • Asking wh-questions to the student rather than multiple-choice (P1-1)

Fill-in-the-blanks is the item most frequently preferred by the junior teacher trainees. This type of item can involve the options that will be chosen or require the learners to write the appropriate structures according to the given context. In the scenarios, P1-2 gave the answer that “I would evaluate their grammar knowledge by expecting them to fill in the blanks in the sentence with the correct grammatical structure”. Although the scenario was related to the assessment of the listening skills of young learners, he did not mention the use of listening for assessing grammar. Therefore, he wants to conduct an isolated assessment focusing only on the grammar knowledge of the students. Moreover, P1-2 accentuates that he will administer written exams involving items focusing explicitly on grammar.

P1-1 also has the thought that it is more appropriate to assess the lower secondary school students with a grammar exam because he will have clearer results, as explained below.

In lower secondary school, the child gradually learns the things he needs to learn in this grammar. For example, s/he learns "can", s/he learns "should". Or s/he starts to learn tenses slowly. For example, in my opinion, the easiest thing for that age group and the thing they need to learn the most is actually simple present tense. You can do direct grammar exams on it now. In my opinion, you can't really measure this by asking in an indirect way. Since this age group has learned much more grammar topics, it is more efficient to ask them directly with a classical exam rather than piece by piece in order to see how much they have learned and how much they can use the rules (Semi-structured Interview III).

In a similar vein, P1-3 and P1-4 also noted that they would give an exam consisting of fill-in-the-blanks items to their students. Regarding this, P1-4 said that she was not sure whether the primary school students were taught structures, and she added:

If they study grammar, I think I want to measure it with a test in order not to force it too much or because I think it will be better at that level, you know, at the primary school level. [What kind of a test can that be?] I mean, there are fill-in-the-blank tests, you know, like what comes here. Or with a test paper where the visuals, for example, the pictures, exist more. I mean, at least I think it may be more evocative. Yes, I will do that (Semi-structured Interview III).

Furthermore, P1-4 stated that she would give assignments to her lower secondary school students every two or three weeks. These assignments would consist of three or four pages, focus only on grammar, and involve various types of items such as fill-in-the-blanks and multiple-choice questions.

Stating in the interview that grammar should be assessed explicitly until the upper secondary school, P2-1 wrote in the scenarios that the grammar assessment would not be carried out if there was no new structure taught. However, if any new structure was taught to the primary school learners, the assessment would be done with the fill-in-the-blanks items.

Another participant, who would like to employ these items, P2-2, stated “A sentence is given, a blank space is left, and it is asked to fill in the blanks according to the grammar of the sentence”. He adds that it should not be multiple-choice and that “[he will] give parentheses; let the child create the sentence himself or herself. So [he will not] give the child a ready answer. Let the child create it”.

The junior pre-service English teachers also prefer using multiple-choice questions with the aim of assessing the grammar knowledge of the young language learners. Of these teacher candidates, P1-2 expressed his views on this issue as follows:

We can ask them to find the correct grammatical structure among these choices that also involve the wrong ones. But I don't think it is possible at the primary school level. It is not possible, at least until lower secondary school. That's my personal opinion. I can explain the reason for that as follows: Because the level of English learned in primary school is a bit low, grammar is not completely learned, and certain structures are used all the time. As we have also learned, 5 tense structures and 4 tense structures are considered sufficient. However, this is not the case, of course, but when a student thinks in Turkish and tries to explain something, he/she cannot find the right grammatical structure. The reason for this is actually not the student, but the structure of the education we provide. Therefore, it is not correct to evaluate the grammar of that child. It may be unethical (Semi-structured Interview III).

P1-3 was of the opinion that the students do not have too much knowledge about grammar and simple structures such as simple present and simple continuous tenses are taught of them. Therefore, she thinks that she should not ask difficult questions to her students. In this regard, she wants to employ multiple-choice items for which the students will not have too much difficulty finding the correct answer.

P2-4 stated that she would not assess grammar by asking the second graders to speak or write in the target language because of their low English level. For this reason, she would resort to traditional items such as multiple-choice questions and true-false statements. Of these, the

true-false statements would require the students to write “true” or “false” according to the grammatical accuracy of the given sentence.

In contrast to these practices, which involve the explicit assessment of grammar and are called weak literacy indicators, the practices enabling the implicit assessment and labeled strong literacy indicators will also be implemented by the junior pre-service English teachers. These practices consist of the implicit grammar assessment done within the language skills. For example, assessing grammar in speaking is one of them, and the teacher trainees emphasize it in various ways. With regard to this, P1-2 and P2-3 said they would assess how their students used grammar in a speech to introduce their family members. Likewise, P1-3 and P1-4 also “assess from the sentences they made when they talked” although items such as multiple-choice and fill-in-the-blanks would also be employed by them. Moreover, it is to be noted that when asked whether she would prefer an implicit or explicit assessment of the language components, P1-3 commented as follows:

I prefer to assess alongside language skills. We can see the words there, whether it is an essay or a paragraph written by the student. Grammar as well. We can see from there how s/he makes sentences. I mean, we can also understand the pronunciation when speaking. There is no need to assess them separately (Semi-structured Interview III).

In the scenarios, P2-2 wrote that he would ask questions to his students about the content presented in the audio. These questions would ask “whether they had the object they just heard in the listening sample and ask where it was in their house, what color it was, and whether they liked it or not”. In addition, oral communication tasks would also be employed for assessing both speaking skills and the use of grammar. Emphasizing the importance of interaction between the teacher and student, he expressed it as follows:

In order to assess language components, I would observe the components in natural oral communication tasks. Without interfering with the communication, I would apply “recasting” if the student made a mistake. Communicative competence is where the general understanding can be revealed, so I would create the said interactive tasks and evaluate the language components through these tasks (Scenarios for Foreign Language Assessment).

Role-playing would also be a practice for implicit assessment of grammar. Related to this, P2-3 wrote in the scenarios that the students would be separated into pairs or groups of three, and they would act out by using simple permission and request statements. Through this practice, she would assess both speaking and grammar.

P2-4 accentuated that she would assess the grammar used in the dialogs between the peers in the 3rd and the 4th grades. She stated “I can ask them to create dialogs among their friends with simple sentences, and in this way I can assess their grammatical competence”.

Moreover, for the lower secondary school students, she would use the question and answer as an assessment practice in which the learners would answer the questions about the listening text asked by the teacher. Thanks to this practice, the learners would produce sentences in English, and the teacher would assess how they used grammar in their speech.

Assessing grammar in writing is another frequently mentioned practice for grammar assessment. With regard to this, P1-1 would draw on the pieces of writing written for giving directions according to a sketch or describing a picture. Additionally, the short paragraphs the learners would write with the aim of introducing their family members would also give the necessary data for grammar assessment.

The grammar used in the sentences the students will write to answer the open-ended questions asked about the listening text will be focus of P1-3. For this, she wrote “When they answer the questions, we can assess them according to the vocabulary and grammar they use”. In a similar vein, the students’ written answers will be taken into consideration by P2-2 while assessing grammar. In relation to this, he said in the interview:

“In early-age learners, grammar is more often manifested in written performance. So in the lower secondary school classes, for example, it doesn't matter whether it's a reading exam or a simple open-ended question related to listening. Their grammar can be assessed on the answers they give to them” (Semi-structured Interview III).

As another context where the students would use grammar, P2-1 mentioned the scripts written by the small groups of students for acting out later. These scripts would be used as materials for writing and grammar assessments. However, it is important to note that this will be a group assessment on grammar.

P2-3, P2-4, and P2-5 mentioned that the short stories written by the students would be the focus for grammar assessment. Besides, P2-3 expressed that “it would be better if the assessment was carried out in a process rather than once or twice on certain dates”. As the reason for this thought, she said “I believe that an implicit assessment can give a more accurate result”. However, it should be noted that she also wants to carry out an explicit assessment of grammar.

Furthermore, P2-4 noted in the Opinion Journal IV that her students would write dialogs for acting out, and she would assess the grammar used in these dialogs. As compatible with this, when asked whether she prefers an implicit or explicit assessment of grammar, she expressed her opinions with these sentences:

Adapting it to some reading or listening did not seem that easy to me or with any skill unless we integrated it with speaking or writing. Therefore, if I wanted to measure students' grammar, I would use an assessment method in which I integrate grammar with other skills, with an emphasis on speaking and writing. And as I just said, rather than giving it directly to the student, I would implicitly present it to the student (Semi-structured Interview III).

According to P2-5, diaries are another type that the learners will write and the teacher will use for his assessments on grammar. Regarding these assessments, he emphasized that the assessment should not strictly focus on grammar in such a way that the students would get scores according to their grammar errors. Instead of that, the focus should be on the students' development in grammar.

Vocabulary is another language component in which any language learner should have enough competency to use the target language. The practices that the junior pre-service teachers plan to employ regarding this component fall into two categories: weak literacy indicators and strong literacy indicators. The first of these means the explicit assessment of vocabulary knowledge through classical testing items such as fill-in-the-blanks, multiple-choice questions, and matching. In contrast to this, the strong literacy indicators involve the implicit assessment of vocabulary use with language skills. To be more precise, the teacher focusing on whether the learner is able to use various words in an appropriate way in his or her speech or writing is an indicator of strong literacy. Additionally, a teacher considering whether the learner is able to comprehend a reading text or audio by drawing on the words existing in these materials is another practice that can be classified under the category of strong literacy.

Table 18

Assessing Vocabulary: Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Multiple-choice questions (P1-2, P1-4, P2-1, P2-2) • Fill-in-the-blanks (P1-1, P2-1, P2-2) • Asking the synonyms of the words (P1-4, P2-3) • Matching (P2-1, P2-2) • Assessing whether the students remember the words (P1-2) • Language components: Inessential at this level (P2-5)	• Assessing vocabulary in speaking (P1-1, P1-2, P1-3, P1-4, P2-3, P2-1, P2-2, P2-3, P2-4) • Assessing vocabulary in writing (P1-2, P1-3, P2-1, P2-2, P2-3, P2-4, P2-5) • Games (P2-3, P2-4) • Asking the student to show the fruit after listening to the teacher (P1-1) • Asking the student to tell the name of the fruit (P1-1) • Assessing listening and vocabulary knowledge together (P1-2) • Assessing vocabulary with reading (P1-2) • Asking the student to tell the word after the teacher gives the first letter (P1-3) • Assessing vocabulary in student's answers to questions (P1-3) • Students writing the appropriate location for the activity (P2-2) • Asking the students which words they heard in the audio (P2-2) • Asking the students to name the activities given in the images (P2-2) • Odd one out question (P2-2) • Students choosing an activity not compatible with the location and explaining why (P2-2)

As one of the constituents of weak literacy indicators, fill-in-the-blanks are the items that will be used for assessing vocabulary. P1-1 is a teacher trainee who will give them to his students. In the interview, he expressed that the vocabulary assessment would be conducted explicitly with an exam including fill-in-the-blank items after the implicit assessment.

Let me discuss primary school. In primary school, it was enough for me to ask them to associate words with each other. I want them to be able to speak the word directly in their daily lives and to visualize what the word they hear is. And I would also like to prepare a small, fun quiz to measure what I saw after I made this association. And I would also like to prepare a small, fun quiz after they make this association. I mean, it could be a test. You wrote the words above. For example, if it asks fill-in-the-blank questions, but of course students at that age may not remember it completely. Because they start to learn very different things in classes like fourth grade and third grade. [...] It is more logical for me to ask them to visualize in their minds in both of these directions and then to see whether they can reinforce this with an exam involving a gap filling (Semi-structured Interview III).

P2-1 also expressed that he would employ fill-in-the-blank items for both the primary and lower secondary school students. Referring to the fact that the learners should have lexical knowledge for using the target language, he commented that the multiple-choice questions would not be appropriate for assessing this knowledge.

If we consider vocabulary and vocabulary use for young learners, we would use fill in the blanks when measuring vocabulary. We didn't use them in grammar. They need to have at least some vocabulary knowledge to be able to use the language. So, there would be fill-in-the-blanks activities. There wouldn't be multiple-choice questions. I don't think it can assess vocabulary knowledge in a good way, in an active way (Semi-structured Interview III).

In an exam that will be prepared “based on the topics or themes covered”, P2-2 will provide a written context with gaps for his students in the lower secondary school. Having the belief that fill-in-the-blanks items will be “healthy” for assessing vocabulary knowledge, he will ask the learners to fill in the gaps with appropriate words by choosing among the given ones.

Multiple-choice questions will be another type of item that will be used to assess the young learners’ vocabulary knowledge. Requiring the learners to choose the correct option according to the stem of the question, these items do not allow the learners to use the words in question, and thus the teachers cannot carry out an authentic assessment of vocabulary acquisition. P1-2 and P1-4 are the teacher trainees who want to employ the multiple-choice items in an exam, which P1-2 especially mentions frequently in the scenarios. They wrote that the vocabulary knowledge of their students would be assessed with this exam, which would involve the words that the learners were expected to know.

Matching a word with a definition or a picture is another method falling into the category of weak literacy. Of the teacher trainees, P2-1 and P2-2 mentions that their assessments would include matching items for detecting whether their students knew the definition of the words in question. In this vein, P2-1 will prepare and administer two forms of matching items. The first of them requires the learners to match the words with the pictures while the second asks them to match the words with their “simple definition or colloquial meaning”. In a similar way, P2-2 expressed that “I believe that vocabulary knowledge should be supported with realia at an early age and that matching it with visuals and a visually supported vocabulary knowledge assessment will be healthy”. Also, matching the definitions with the words that are selected in consideration of the ones covered in the lessons will be in the exams P2-2 will administer.

Two teacher candidates will also use the method of asking the students to tell the words’ synonyms. With regard to this, P1-4 stated she would do this during the lessons as an informal assessment in order to monitor and develop her students’ vocabulary learning. In a slightly different way, P2-3 wrote in her journal, “I prepare one-word cards and collect them in a bag. I can ask my students to choose cards at random and say the synonyms of the words”.

The junior PSETs also mentioned additional implicit and authentic methods in addition to these traditional and explicit assessment practices. Assessing vocabulary use with productive

language skills, namely speaking and writing, is one of them. Assessing vocabulary in speaking was a method that almost all teacher trainees, except P2-5, would like to use. Regarding this, P1-1 noted in the scenarios that he would focus on the words the learners used in their conversation, as explained in the quotation below.

The assessment of vocabulary can be done in many different ways in comparison to grammar or pronunciation. For example, during an impromptu conversation, does s/he use the words s/he uses in the right place and for the right purpose? How are their word choices? etc (Scenarios for Foreign Language Assessment).

P1-2 remarked in the semi-structured interview that “if we directly assess the word, if we ask questions like ‘Do you know the meaning of this word?’ or ‘Do you know the spelling of this word?’, it means rote learning, and I don’t think it’s true”. In line with this view, he plans to assess the vocabulary use in a speech that his students will make for introducing their family members or in a dialog with their peers. However, it must be noted that despite his argument for the implicit assessment of vocabulary, he also wants to explicitly assess vocabulary with a written exam.

P2-1 emphasized in both the interview and the scenarios that she would assess vocabulary in the speech her students would produce during the question and answer activity and while introducing their family members. For the primary school students, he explained how the assessment would be:

Students are expected to correctly use the vocabulary expected to be used when introducing and describing family members. [...] The vocabulary expected to be used would have been taught within the concept of family before the assessment process, and students would therefore have been able to demonstrate this vocabulary knowledge to a certain extent in the assessment process. In order not to restrict speaking skills at this level, students are not informed about the words they are asked to use unless they have to (Scenarios for Foreign Language Assessment).

As for the lower secondary school students, P2-1 said that he would ask his students questions that would enable them to use the target words. Here, the aim is to “see if s/he uses that word or not”. Likewise, P2-2 will ask questions in order to allow the primary school learners to speak and use the words in question. Moreover, these students will introduce their family members and compare them with their peer’s, and the teacher will focus on the lexemes used in their speech.

The students’ words in the speech that will be made for introducing their family members will be paid attention to by P2-3, too. In addition, she will assess her students’ vocabulary acquisition by focusing on “the short narratives describing activities they like or dislike and daily routines”. These narratives will be made by “using the related words in the listening

text” they listen to before speaking. She also mentioned in the scenarios that the role-plays involving the students asking and giving permission with short and simple sentences would provide the context for the vocabulary assessment. As another method for implicit assessment, P2-3 will give the learners vocabulary cards and ask them to make sentences by using the given words.

Card games are both fun and interesting. I can make use of them at all levels. The level of the words will vary, but I can make use of them at all levels. A word is written on a card, and the student is given that word, i.e., randomly selected, and asked to use it in a sentence. Or they can be asked to use it depending on a theme. If the student is really able to do it, depending on the theme, depending on the content, or if they can use that word, they will be able to use it in their daily life in that situation. That's the whole point of learning vocabulary (Semi-structured Interview III).

P2-4 said that “I want to be such an English teacher that my student can use whatever I teach him or her in the lesson”. As compatible with this wish, she plans to “carry out an assessment to assess the vocabulary acquisition of the students in line with the words [she] taught them”. With this aim in mind, like in P2-3, she would also like to use dialogs among the students for assessing vocabulary and speaking. Moreover, drama is another method that she wants to adopt for these assessments.

Assessing to what extent the learners are able to use words in written texts is another practice of implicit assessment of vocabulary. Again, eight out of the nine junior PSETs would like to focus on the use of lexemes in writing. Of these teacher candidates, P1-2 commented that “we can more easily see in writing the quality of the student’s vocabulary knowledge and use”. For this reason, he prefers to assess the lower secondary school students’ vocabulary use in the texts that they will write.

Paragraphs and the answers the learners will write for answering the open-ended questions about audio will be taken into consideration by P1-3 for vocabulary assessments. While doing this, she will pay attention to the variety of words used in these texts. Related to this, she gives the following example:

If s/he uses the same word, other words also have synonyms, or if s/he doesn't use a synonym of a word. For example, if s/he uses ‘important’ for the first time and uses ‘important’ every time, she may not know the other synonyms. For example, ‘significant’ can be used. There are other synonyms as well (Semi-structured Interview III).

P2-1 and P2-3 will assess vocabulary in the answers the students will write for answering the open-ended questions based on the reading or listening text. Also, the scripts of the conversation written by the individual students or the pairs will provide the context for the

vocabulary assessment done by P2-1 and P2-4 while P2-5 will pay attention to the words used in the diaries.

Assessing vocabulary implicitly through listening was also mentioned by P1-1 and P2-2. Of these, P1-1 said he would conduct an assessment in which he would say the name of the fruit to the student and expect him or her to point to the fruit. On the other hand, P2-2 would ask the primary school students to identify activities, objects, and the colors of these objects mentioned in the audio and to create their “original sentences” with these words.

Games are another way of vocabulary assessment mentioned by one of the junior PSETs. P2-3 expressed that she would prefer card games like taboo instead of quizzes. She gave the following explanation related to this game:

Instead of quizzes, I think it is more permanent to carry out vocabulary assessment with games. I mentioned this in my diary. A game like Taboo. A spare 4, 5 forbidden words for a word would be much more useful for a student with visual intelligence. The student should learn both one word, the related word that he/she should not use, and the expressions of other words that can express that word. I think these can be more permanent (Semi-structured Interview III).

Assessing the Language Components: Adolescent Language Learners

The junior PSETs plan to use various methods for assessing the adolescent language learners’ knowledge about and use of language components, as in the case of the young learners of English. Therefore, it is again possible to categorize these methods under the categories of weak and strong literacy. As in the previous section, weak literacy involves methods or items that allow only the explicit assessment of language components in isolation. On the other hand, strong literacy consists of methods appropriate to the implicit and authentic assessment of the language components through the four language skills. In consideration of this categorization, the tools that the PSETs would like to employ for explicit assessment of vocabulary are fill-in-the-blanks and multiple-choice questions.

Table 19

Assessing Grammar: Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the blanks (P1-2, P2-1, P1-4) • Multiple-choice questions (P1-2, P1-4)	• Assessing grammar through speaking (P1-1, P1-2, P1-3, P2-2, P2-3, P2-4) • Assessing grammar through writing (P1-2, P1-3, P2-1, P2-2) • Assessing grammar through reading (P1-1)

P1-2 mentioned that the multiple-choice questions could be used for assessing grammar. During the interview, he expressed that these items would be more appropriate for adolescent learners. Moreover, he would assess these learners' grammar through the fill-in-the blank items in an exam. P2-1 will also ask his students to fill in the gaps with appropriate structures after listening to the audio. As different from these PSETs, P1-4 will give her students grammar assignments in which there will be multiple-choice and fill-in-the blank items every two or three weeks.

The junior PSETs also want to assess grammar through speaking and writing. For assessing grammar through speaking, the PSETs will carry out various practices. Of these teacher trainees, P1-2 wrote that the grammar used by the learners for acting out a story and making a short speech about past habits and traditions would be paid attention to. Additionally, he emphasized in the interview that "it is more correct to assess grammar in conversation". P2-3 also noted that she would assess the grammar used in the speech about past habits and traditions.

As another student teacher who will conduct implicit grammar assessment, P2-2 expressed his ideas as below:

At higher levels, I would expect the grammatical structures to be established in speaking skills as well. In other words, in the early ages, it should be examined more in terms of written performance, but in the later levels, I expect to see it in that as speaking skills come to the fore. For example, when I compare pronunciation and grammar, in a speaking assignment or performance study in high school, a child can cause many misunderstandings or deficiencies due to lack of grammar, while pronunciation can only impair comprehensibility to a point, unless it is very bad. Therefore, grammar is more prominent than the other 2 groups, especially at advanced levels in terms of speaking compared to pronunciation (Semi-structured Interview III).

Debates and presentations will be the tasks that P1-3 will use for assessing both speaking skills and grammar. Related to these assessments, P1-3 stated that she would prefer to record the students' speech while carrying out the tasks for more robust assessments.

Regarding the implicit assessment of grammar, P2-1 pointed out that he would prefer to assess grammar through language skills, especially in upper secondary school. He also emphasized that some specific topics could be used to enable the learners to make sentences with the target modals and tenses, as given below.

By the time we reach upper secondary school, I don't think there will be any need to assess grammar explicitly. In this process, I would not make an additional evaluation since they will have already learned enough grammar to be able to use the language competently enough and to use it in their daily lives. Again, it would be within the skills of using grammar. Listening, writing, and so on are not all included until lower secondary school. For example, in the lower secondary school, I would never teach it in A1. In A2, I would evaluate it in writing. For the upper secondary school, they spread across all activities. It would be spread over all language skills. I would evaluate grammar in listening, writing, reading, and speaking. [Can you give an example?] I think these can be integrated very easily. Because when I say speaking, for example, if we are assessing the use of tenses, for example, I said to the student, "Do you think we will be able to go to Mars in the future?" There is no way that the child cannot use the future tense here. By using the past, he can finally say that we did this in the past. He can say that so far, we have done this. Here, s/he can use all the tenses, including present and future tenses. I'm giving examples based on tense, but other than that, topical tense usage can affect the use of tense in a comfortable way, plus modal usage. Students will not speak with certainty because it will happen in the future. Topics that affect the use of modals and tenses. Apart from that, if I wanted to specifically measure the use of pronouns, conjunctions, articles, etc., I had to make them explicit. Because the topic may not be able to direct them either (Semi-structured Interview III).

Question and answer is one of the methods that P2-4 mentions for grammar assessment. It will take place after listening to audio, and the learners will get feedback on their errors. In her Opinion Journal, she explained what kind of methods she would prefer for assessing grammar.

I would do an assessment of grammar not with the usual methods such as multiple-choice questions or true-false questions, but with a form of assessment that would both make the process more fun and require the active participation of my students (Opinion Journal IV).

Assessing grammar through writing is another type of implicit assessment that the junior teacher trainees mentioned for adolescent language learners. In the interview, P1-2 said that grammar could be implicitly assessed in writing by considering such questions as: "Are the grammatical structures used in the student's writing correct or are there errors?" In the scenarios, he noted that they would write texts about the given topics in the classroom within the allocated time.

P1-3 mentioned paragraphs, essays, summaries, and answers to the open-ended questions as the contexts for grammar assessments. The learners will write the paragraphs and essays

about a topic they choose or a topic given by the teacher while the summaries will be written about a text giving information about a city or a neighborhood. As for the answers to the open-ended questions, they will be the answers written to the questions asked about audio.

Like P1-3, P2-1 will also pay attention to the grammar used in the answers to the open-ended questions asked after listening to the audio. In addition, the texts written by the learners will be focus of the assessment he will conduct. In relation to this, he wrote “I would assess grammar in the context of the written texts and take into account even the smallest grammatical mistakes (syntactic and morphological mistakes)”. For the sentences that will be written for comparing the cities, he noted that “the students' use of grammatical structures for comparison will be specifically checked, and any other usage will be noted for later correction”.

The information required for the grammar assessments that P2-2 will conduct will come from the responses to the open-ended and "opinion-based questions." These questions will be asked about a reading text in a written exam, and the learners are expected to give answers by drawing on the reading text and their own opinions concerning the issues mentioned in the text. In the interview, he explained what his assessments would be:

It should be implicit in terms of assessment. And the main goal should not be complete grammatical accuracy at an early age or at the A1 or A2 level. However, this criterion should still be sought in assessments that do not directly aim to assess grammar at advanced levels. So in a reading comprehension question, the answer should be grammatically correct, especially if it is open-ended (Semi-structured Interview III).

In contrast to the assessment of grammar through productive skills, the grammar assessment through reading was also mentioned by one of the PSETs, P1-1. Regarding this, he wrote in the scenarios that he would conduct an assessment on whether the students were able to order the scrambled sentences by drawing on conjunctions and tenses.

As in the grammar assessment, it is possible to categorize the methods mentioned by the junior PSETs for assessing vocabulary as indicators of weak and strong literacy. The methods labeled weak literacy indicators are the traditional methods, which are far from real-life practices. On the other hand, the methods labeled strong literacy indicators are the ones that allow the implicit assessment of vocabulary through language skills and are compatible with real-life situations.

Table 20

Assessing Vocabulary: Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the blanks (P1-2, P2-1, P2-2) • Multiple-choice questions (P2-1, P2-2, P2-3) • Writing definitions of words (P2-1, P2-3) • Picking a card out of the pocket and saying the synonym of the word (P2-3)	Assessing vocabulary through listening (P2-2) • Students choosing an activity not compatible with the location and explaining why (P2-2) • Odd one out question (P2-2) Assessing vocabulary through speaking (P1-1, P1-2, P1-3, P2-1, P2-2, P2-3, P2-4) • Assessing vocabulary in a short speech (P1-1, P1-2, P1-3, P2-1, P2-2, P2-3) • Assessing vocabulary through role-play (P1-2, P2-1, P2-2, P2-4) • Assessing vocabulary through games (P2-3, P2-4) • Asking students to describe different aspects of words (e.g., zoo, theater) (P2-1) • Assessing by singing songs (P2-4) • Question and answer (P2-4) Assessing vocabulary through writing (P1-2, P1-3, P2-1, P2-2, P2-5) • Open-ended question (Reading/Listening and Writing-Integrated) (P1-3, P2-1, P2-2) • Assessing vocabulary in a summary of text written by the student (P1-3) • Writing a paragraph or a text by using the given 10 or 20 words (P2-1) • Writing a new sentence with the given word (P2-2) • Students writing the appropriate location for the activity (P2-2)

One of the items that the teacher trainees will use for explicit vocabulary assessment is fill-in-the-blanks. In a written exam, P1-2 will give his students these items, which will be prepared based on a reading text, and ask them to fill in the gaps with the appropriate words

by reading the given text. On the other hand, in the assessments that will be done by P2-1, the fill-in-the-blanks items will be answered after listening to audio. Regarding this, he said,

For listening, the vocabulary can already be easily integrated into the listening. Even if there is no such listening text, I, as a teacher, can create one. I can do a role-play by myself. Or with a friend. Or I can talk by myself like this. Through this, through listening, comprehension and filling in the blanks can be asked (Semi-structured Interview III).

In the Opinion Journal, P2-2 wrote he would employ both fill-in-the-blanks and multiple-choice items in the formal assessments. However, he stressed that these items should constitute a small part of the assessment in comparison to the use of words through productive language skills, and he explained the reason for this as below:

Assessment of vocabulary should be embedded in the tests with scores. To expand this further, I believe vocabulary is an aspect of the language that can lead to confusion, especially when put together with a collection of words. Due to this, I would immerse vocabulary in the previous four skills' activities. Certainly, vocabulary-based tests such as fill-in-the-blanks and multiple-choice tests can be applied, but they should take place after the student's performance (e.g., after allowing the students to discuss a text, write about the text, etc.). When applied to actual tests (that involve scoring), fill-in-the-blank and multiple-choice tests can be referred to, but they should not involve the majority of the test. Actual production means not to memorizing the meanings of vocabulary but learning how to use them in a felicitous way (Opinion Journal IV).

Multiple-choice questions will also be included in the vocabulary assessments. For administering these items, the teacher candidates will give them to the learners as reading text or only words as prompts. P2-1 stated that the multiple-choice questions in his assessments would be asked based on a reading text, and in order to reduce guesswork, the questions would require the students to make inferences by drawing on the key words in the text. As different from this, P2-2 will assess whether the learners have knowledge about the synonyms of the words. For this, he will give the words and ask the students to choose the synonym of the word among the options.

The teacher trainees also mentioned other methods compatible with the explicit vocabulary assessment. Another one that two junior PSETs will use is asking the students to write the definitions of words. Of these, P2-1 wrote in the scenarios "Students will be given words related to different environments and asked to describe their different characteristics in the space below the relevant word (birthday party, zoo, theater, etc.)". For this, P2-3 will give vocabulary quizzes because she thinks that "it will also be good to make use of vocabulary quizzes at regular intervals". In these quizzes, she will ask the learners to write down the meanings of words. Although she prefers administering quizzes for assessing vocabulary knowledge, she notes that she thinks "quizzes are a bit of a temporary result". And she adds, "It can be applied again. But it would not be right to take it as a single assessment. I think

more than one method can be used”. Moreover, P2-3 will ask adolescent students to pick a card out of the box and to say the synonym of the word they chose randomly. However, these methods cause rote memorization rather than the acquisition of vocabulary, which is an obvious common feature of them.

In contrast to these traditional and explicit assessment methods, the junior PSETs will also implicitly assess vocabulary through listening, speaking, reading, and writing skills. Implicit assessment of vocabulary through listening involves the learners making inferences by drawing on the words in the audio and using the target language by adopting these inferences and their own opinions. In line with this type of assessment, P1-2 emphasized both in the interview and in the opinion journal that he would prefer assessing vocabulary through questions requiring the assessees to have knowledge about or guess the meanings of some important words in the audio.

In the scenarios, P2-2 wrote that he would design a listening exam that involved audio about a group of people making their holiday plans and such words as swimming, beach, climbing, canoeing, trekking, and so on. In this vein, he would carry out such an assessment of vocabulary as:

In order to analyze grammar and vocabulary knowledge, an additional part can be added. Using the activities and locations from the sample, an “odd one out” question can be added that presents a location and various activities attached to it. Students can choose the one that does not fit the location and explain why. Additionally, they are also required to write the suitable location for that activity. Consequently, this will not only test their vocabulary but will also allow them to present their ability to make new meaningful connections as a written answer (Scenarios for Foreign Language Assessment).

Assessing vocabulary through speaking involves the teachers focusing on how much the learners are able to use the words appropriately while making a speech individually or interacting with their teacher or peers in English. P1-1 would like to do such an assessment when he becomes an English teacher after graduating from the ELT program. He expressed his opinion about this with these sentences:

It is much more important for me that the child [...] is able to choose the words he/she wants while speaking fluently, compared to taking them one by one. [...] It doesn't make much sense to me to consider words piece by piece. Things he can memorize. I want him to learn it, and I think it would be much more logical and productive for the student if I handled it in an implicit way, that is, through something that he could talk about face-to-face (Semi-structured Interview III).

As compatible with this idea, he wrote in the scenarios that “I would look at whether s/he was making the right choice of vocabulary [...] when presenting the work s/he analyzed and

the biographical sequence orally”. In addition, he exemplified how he would expect the learners to use more formal words in a context of interaction with a boss or a customer, as given in the excerpt below.

During a profession, while practicing a profession and talking to a foreign boss or a foreign customer, you are actually looking at how they explain the right word, the word they want, and the word they need there. Here, you expect a slightly more academic, let's not say professional, word choice from the student (Semi-structured Interview III).

The vocabulary used by the adolescent language learners in short speeches like presentations and role-plays will be the focus of assessments by P2-1 and P2-2. In these oral products, P2-1 will expect the learners to use words and expressions taught in the lessons, and he will do his assessments by watching the recordings of these performances. On the other hand, P2-2 will pay attention to the “lexical inventory” that his students will use in these conversations and their ability to use these words in a contextually appropriate way.

Individual talks about customs and past habits will also provide context for the vocabulary assessments by P2-3. She noted that the learners would be informed in advance that their vocabulary use would be assessed. Besides, she added “I want my students to do self-assessment and assess themselves before I do my own assessment”.

Card games will also be employed by P2-3 as a part of informal vocabulary assessments. Taboo is one of these games, and it involves the learners as players trying to enable their peers to guess the target word while playing it. In this way, both the students explaining the word and the peers trying to find the word in question will use their vocabulary knowledge and the target language. In addition, P2-3 explained another way for vocabulary assessment, which she called a card game.

A word is written on a card, and the student is given that word, i.e., randomly selected, and asked to use it in a sentence. Or they can be asked to use it depending on a theme. If the student can really succeed, depending on the theme, depending on that content, or if he/she can use that word, s/he will be able to use it in that situation in his or her daily life. This is the purpose and intention of vocabulary learning (Semi-structured Interview III).

Question and answer is a method P2-4 will draw on for vocabulary assessments. The students will express their answers to the questions asked about the audio they listen to in advance. She will pay attention to how successfully the students are able to use the words in their answers. Moreover, she emphasizes that she will provide feedback on the students’ lexical errors in these answers.

The texts written by the students for various purposes will also provide the teacher candidates with data for vocabulary assessments. In this regard, P1-2 wrote that he would assess vocabulary use in the e-mails or complaint letters the adolescent language learners wrote. On the other hand, the words in the answers written to the open-ended questions about the listening text will be the focus of P1-3, P2-1 and P2-2. Moreover, P1-3 mentioned the summary the students would write after reading a text about past habits and experiences.

P2-1 also emphasized the students being able to understand the words in the reading text and to answer the open-ended questions concerning the text by drawing on their inferences, as given in the quotation below:

In reading, there would be open-ended questions under the paragraph. He would also write at the same time. The key words needed to answer these questions would be the words in the paragraph that he could also understand from the context. The aim here is both to understand the word when he sees it in the paragraph and to measure whether he knows it directly or not. Open-ended questions are also needed for this reason (Semi-structured Interview III).

Additionally, the two PSETs, P2-1 and P2-2, will give their students some words and ask them to write sentences or pieces of writing using the given words. With regard to this, P2-1 said “They can be asked to write a text within a certain topic, giving a certain 10 words or 20 words, giving some target words”. Likewise, P2-2 pointed out that the learners would be proficient enough to use the target language in a written way, and thus they could make original sentences with the given words:

A completely new, separate sentence with that word can be asked to be created, especially in upper secondary school. I now assume that the child has used it communicatively. Therefore, I think it will not be difficult for him or her to display this in the written performance (Semi-structured Interview III).

P2-2 will also assess the use of vocabulary in answers to open-ended and opinion-based questions. These questions will be based on a reading text in a written exam and require the learners to understand the words in the text and use English for answering them.

Scoring, Interpreting, and Using the Assessment Results

In this research, scoring means the procedure of assigning scores to the assessment products, including the answers to open- or closed-ended questions and the performance displayed through productive language skills. Taking this definition into consideration, the scoring practices that the junior PSETs were planning to employ were categorized into the categories of weak and strong literacy. The scoring procedures falling into the category of weak literacy

involve practices in which the teacher, as a rater, does not give scores according to a specific answer key, checklist, or rubric. That is to say, the rater draws on his or her own personal judgments without taking any predetermined criteria into consideration or assigning scores according to the criteria, which are not well defined or address a limited part of the structure to be assessed.

Table 21

Scoring the Answers or Performances

WEAK LITERACY	STRONG LITERACY
<i>Scoring according to Personal Judgment</i>	<i>Scoring according to a Rubric or Table of Criteria</i>
• Scoring according to the criteria in mind (P1-1) • Scoring by focusing on both strengths and weaknesses (P1-4) • Scoring strictly with little flexibility (P1-4) • Unable to define the criteria for scoring writing (P2-2) • Scoring according to observation or by using her initiative (P2-3) • Scoring with a more tolerant attitude (Young learners) (P2-4) • Distributing equal scores for each component (P2-4)	• Scoring according to a rubric (P1-1, P1-2, P1-4, P2-1, P2-4) • Scoring according to a table of criteria (P1-2, P2-4)
<i>Scoring according to Vague Criteria</i>	<i>Scoring according to an Answer Key (P1-2, P1-3, P2-2, P2-3)</i>
• Scoring according to whether the student is able to understand the whole text (P1-1) • Scoring according to how many times s/he can understand the question (P1-1) • Scoring according to whether s/he is able to ask questions about the listening text (P1-1) • Scoring according to errors (P1-3) • Scoring based on the students' comprehension of the information in the text (P1-3)	<i>Scoring according to Specific Criteria</i> • Intelligibility (P1-1) • Fluency (P1-1) • Word choice (P1-1) • Grammar use (P1-1) • Scoring according to the determined word count (Writing) (P1-3) • Acting out the script in a natural way (Naturalness) (40 points) (P2-1) • Fluency (40 points) (P2-1) • Correct use of grammar and vocabulary (10 points) (P2-1) • Creating a meaningful text (10 points) (P2-1) • Pronunciation (10 points) (P2-1) • Fill-in-the-blanks (30 points) (P2-1) • Open-ended questions (60 points) (P2-1)

-
• Paying attention to each detail without informing the students about it (P1-4) • Scoring with regard to the rate of answering open-ended questions (P2-1) • Scoring to what extent the performance or product is near the goal or objective (P2-3) • Scoring the learning process (P2-5) • Paying attention to individual progress (P2-5) • Paying attention to reading comprehension (P2-5) • Paying attention to how the students can use what they understand (P2-5)	• The value of the main points in the video (40%) (P2-1) • Value of details in the video (60%) (P2-1) • Equal score for each blank (P2-1) • Use of gestures and mimics (P2-1) • Ability to work in harmony with the group (P2-1) • Scoring: Summarizing main points and ordering events accurately (P2-1) • Cohesion (P2-2) • Accurate use of the structures (P2-2) • Answers' relatedness to the question (P2-2) • Intelligibility (P2-2) • Fluency (1-5 points) (P2-2) • Grammar (1-3 points) (P2-2) • Pronunciation (1-3 points) (P2-2) • Vocabulary (1-4 points) (P2-2) • Purposeful communication (1-5 points) (P2-2) • Novelty (0-5 points) (P2-2) • Grammatical range (0-5 points) (P2-2) • Lexical inventory (0-5 points) (P2-2) • Interpreting the information given in the audio (P2-2) • Comprehending the audio (P2-3) • Use of language (P2-5) • Originality (P2-5) • Personal progress (P2-5) • Novelty of language expressions (P2-5) • Communication skills (P2-5)
-

Some answers given by the junior PSETs to the questions asked in the interview and the scenarios indicated that they would draw on their personal judgments while scoring the

students' answers to the questions or products such as writings, speeches, or performances produced as a result of various assessment methods. For example, P1-4 answered in the scenarios: "Since the A2 level is a stage where language awareness is a bit more developed, I expect them to understand 70-80% of the theme of the event and the speech in what they listen to". Instead of specific criteria, she noted she would take account of "whether s/he paid due attention to pronunciation, whether s/he paid attention to the modal or tense that should be used in the subject, and whether s/he was trying to be sufficient in this regard". Besides, P1-4 would score "not only by focusing on the weaknesses of the student but also by taking into account the areas where they are sufficient" without considering predetermined criteria.

The answer written by P2-1 will also indicate that he will not rely on enough specific criteria while awarding scores to the adolescent learners reading comprehension.

I consider the assessment a success if, according to the criteria I have specified, a summary of the main points has been produced and the events in the lives of the people whose biographies are given have been listed correctly. I deduct points for failure to meet these criteria. Apart from these, I do not include any other skills or components in the scoring (Scenarios for Foreign Language Assessment).

Regarding how to score the writings of young and adolescent learners, P2-2 accentuated in the interview that he would not be able to determine specific scoring criteria because of the open-ended nature of the writing.

If I have a group of young students, especially if I have given them the freedom to write, there is no such thing as a score; there can't be such a thing: Write about what you did during your summer vacation. As I said, there would be only one scoring criterion for me: Did he/she give an answer depending on the question? Other than that, minus 1 for every language structure error, and this is not very likely to happen at an early age. I would probably say that the maximum score that can be cut is for language structure errors. I feel like writing measurements and products at an early age cannot depend too much on a fixed score criterion. [...] It could be like this: For example, upper secondary school students could come up with an element of writing. The content of this could be differently articulated statements that address a single truth. For example, an open-ended question based on a reading text is a very good example of this. I mean, I think writing is a very free space in terms of expression, so I can't put a fixed wall around it. Let it address the truth (Semi-structured Interview III).

For the written and oral products, P2-3 pointed out that "I can make a rating using my initiative and observations. I can get these results from my own point of view according to the success and failure rates of the questions".

Unlike her peers, P2-4 wants to assign scores to her prospective students based on the skills they are more successful at. For expressing this, she stated "If I knew my students at certain levels and I knew what the student was really good at, I would probably assess that student in a way that would focus more on that skill but not hinder the others". Also, she exemplified:

“For example, a student is better at speaking. But his or her writing skills are not that developed. [...] Of course, taking these into consideration, I would evaluate that student in a way that he or she could get more points on speaking skills”. Moreover, P2-4 wrote how the adolescents were able to use the language components in their written and oral inferences, comments, and criticisms after reading the text would be scored. However, rather than measuring according to a set of criteria, she would “distribute the points evenly without ranking them in order of importance since each component would be important, especially for a student at B1 or B2 level”.

In addition to drawing on their personal judgments, the junior PSETs will also carry out scoring according to some vague criteria. For instance, in the scenarios P1-1, mentioned three criteria for scoring the listening skills of the young learners. As the first criterion, “whether s/he can understand the whole” of the listening text was written. The second one was “how many times s/he can understand the question asked”, and the third was “whether s/he can ask a question by inferring something from the speech or text s/he listens to”. Although they refer to being able to comprehend the oral texts and to speak in order to ask questions based on the audio, the criteria are not specific enough to give high scores to those who successfully attain these objectives and low scores to those who fail.

While giving scores to the texts written by her adolescent students, P1-3 will also not apply clearly defined criteria. Instead, she will score according to such vague criteria as “how s/he writes and whether s/he expresses herself or himself well”, “the grammar and use of words”, and “self-expression and mistakes”. In a similar way, she wrote for scoring the summaries based on a written text that “when we give points, we can give them according to their compliance with the criteria”. Although these criteria will be related to whether the student is able to analyze the text well and the use of grammar and vocabulary, it is unclear exactly what these criteria will be and how it will be decided whether and how much of the required qualities are present in the text.

The sentences written by P2-1 in the scenarios also revealed that he would not adopt robust criteria while judging the answers of the adolescent learners to the open-ended questions about the listening text. He noted that the section of open-ended questions would be scored out of 60 points, and the points that would be awarded for the answers might vary according

to the difficulty of answering the questions. However, no specific criteria were mentioned concerning the assignment of scores to answers with different qualities.

P2-2 wrote the elements he would consider while scoring the speaking skills of the young learners. Although he mentioned the highest and lowest scores that could be awarded for each element, he did not clearly explain which criteria would be considered for judging the quality of the speech, as given in the quotation below:

The overall scoring would consist of four categories. Fluency (1-5), grammar (1-3), pronunciation (1-3), vocabulary (1-4), purposeful communication (1-5). The majority of the points can be earned depending on how well the children produce utterances and their ability to give correct responses. Vocabulary comes in as the second most important category in order to evaluate students' ability to create distinguished utterances and not fall into repetition. Grammar and pronunciation hold the lowest priority. Pronunciation relies on understandability, not native-likeness. Grammar structures are expected to be performed correctly, yet it gives freedom for minor mistakes such as subject-verb agreement. Mistakes such as "open can I window?", however, are an indication of "errors" so the students have the chance to lose points if they organize the parts of a basic sentence structure (Scenarios for Foreign Language Assessment).

For P2-3, the basic criterion in scoring the written or oral products will be to what extent the performance or product is near the goal or objective in question. In this manner, she will separate the scores into levels according to their proximity to the goal. However, this criterion and the scoring levels are not discrete enough, as in the excerpt below:

I said that young children first learn to introduce themselves. I mean, a student who can more or less introduce himself or herself gets a grade of 80 in my class. A student who can fully explain and express himself or herself gets a grade of 100. But if the words are jumbled together, if something is formed with sentences that are far away from each other, the product is 60. I am saying this for writing as well as for speaking. I try not to keep it too low at first, you know. So, as I said, in order not to be a deterrent, as the level progresses and the expectations increase, I will make the evaluation accordingly. But as I said, I prefer to go targeted. This would be the framework. Just because it is perfect in terms of its proximity to the target does not mean that it is 100. For me, being close to the target makes it more valuable. I make an evaluation accordingly (Semi-structured Interview III).

Drawing attention to the importance of students' individual progress in language learning, P2-5 noted in the scenarios that he would score how they improved over time. While he did not specify enough how he would measure this progress, he mentioned three components for rating the written products: language use, originality, and to what extent s/he internalizes chunks. Acknowledging that deciding how well these criteria were met in the exam would be difficult, he pointed out that individual progress was also important. Also, he added that giving scores to classroom performance was necessary since the students' language learning backgrounds were different and fairness among the students should be ensured.

I score according to personal progress and the level of language learning. Every child's learning conditions, environment, and so on are different. With these differences, it is important to subject

students to the same test and see if they make the same progress (Scenarios for Foreign Language Assessment).

Because if we evaluate language learning as a process, while making this evaluation, we have to see how far the student has come in this process, and as I said, not everyone has the same language skills, and they are not equal. In order to ensure equality, there is actually a point that we call the opinion grade, the point that the teacher gives to the student in a way based on the observations he/she has made. Things like this can come into play. Because someone, for example, has lived abroad. He speaks very well. The other one didn't know anything when he came to school, but now his speaking skills have increased a little bit and he is determined (Semi-structured Interview III).

Rubrics, answer keys, and specific criteria for scoring will also be adopted by the junior PSETs. In the literature of assessment, a rubric is defined as a tool that includes a set of criteria for gauging the quality of a product or performance and assigning scores according to how well these criteria are met in the product or performance. By drawing on this definition, scoring according to a rubric that involved the answers referring to the use of rubrics or a set of criteria emerged as a sub-category of strong literacy.

For example, P1-1 pointed out that he would prepare a rubric for scoring the written products. Mentioning that using the rubric would ensure fairness in scoring, he expressed his ideas as follows:

I need to prepare a scoring scale. The sharpness of the words, language rules, good writing, and punctuation. Whether the word is written correctly or not, and so on. I will score them all equally according to this scoring scale I prepared before. Here, if a student misspelled this word and I gave points to that student, I have to do this to all other students. Because then the students come to me and ask why you gave points to this one and not to me. In my opinion, it has no validity both in the eyes of the students and in my own (Semi-structured Interview III).

In the scenarios, P1-1 wrote that he would also use a rubric to score the oral products produced by young learners and adolescents by taking into account issues such as intelligibility, fluency, word choice, and grammar rules.

As explained in the previous sections, P1-2 has some concerns about the fair scoring of the products and performances emerging as a result of open-ended questions or tasks. Therefore, he mentioned in his journal and the scenarios that “I create a table of criteria and score them in order to make sure that the evaluation is as closed to interpretation as possible”. He will employ these tables while scoring not only the written products but also the oral performances.

P2-1 wrote that he would assign scores according to three main criteria for the speaking performance of the fourth grade students.

I would create a main assessment part consisting of 3 parts: preparing a text containing the desired expressions; performing the text with the group in a natural way (speaking in harmony

with body language and not just by rote); and speaking fluently and with correct pronunciation, and I would score them as 20/40/40, respectively. For the first part, the correct use of grammar and vocabulary will score 10 points, and the ability to prepare a meaningful text using the desired expressions will score another 10 points. Detailed criteria, such as determining the level of fluency in the remaining speaking sections, will use the existing IELTS criteria (Scenarios for Foreign Language Assessment).

Regarding the scoring of speaking performances by adolescents, P2-1 will rate both the enactment of the given story and the speech made about the traditions. To this end, various levels of scores will be determined for each attribute sought in the oral text before the assessment.

I score both parts with 50 points and evaluate both parts under the headings of general fluency, correct usage, use of gestures, and facial expressions. Accordingly, I evaluate the act with 10 points for general fluency, 10 points for correct use, and 20 points for the use of gestures and facial expressions. For the second part, I score 15, 15, and 10 points, respectively. I set the ability to work harmoniously with the group at 20 points (Scenarios for Foreign Language Assessment).

In the scenarios, P2-4 wrote that she would prepare a “scoring criterion” for rating the young learners’ written answers to the questions about the audio. In this criterion, “the language components are a little bit more in the background, and whether they can understand what they read or construct sentences is more in the foreground”. Moreover, she noted that she would inform the learners about the criteria in advance. Likewise, she will design a scoring table for rating the speaking performance of the young learners and adolescents. In this table, grammar, vocabulary, and pronunciation components will be ranked and scored in order of importance.

Answer keys will also be used by the PSETs for scoring the answers to the questions asked about reading and listening texts. P1-2 stated that the scoring procedure in the assessments on reading and listening skills would be easier in comparison to the one in the speaking and writing assessments. As the reason for this, he mentioned the objective nature of the limited response items.

Scoring the products we get from listening and reading can be a bit easier when compared to writing and speaking. Easier again, as I said, I plan to do it with more objective question types, such as multiple-choice questions, such as giving certain words and filling in the blanks for listening. Yes, since the answer key is already clear, it is a fair evaluation compared to writing and speaking (Semi-structured Interview III).

P1-3 also expressed that fill-in-the-blank items could be employed for the listening assessments. Regarding the scoring of these items, she would pay attention to whether the gaps were filled in with the correct answers, and she would use an answer key for awarding scores. Likewise, P2-3 also plans to draw on answer keys for scoring the answers to fill-in-the-blank and true-false items that will be deployed for reading and listening assessments.

Since there will be more specific questions that require clearer answers than speaking and writing, and since the answers will also be oriented towards this, the evaluation will be based on certain points, as you say. What can I do in these situations? I mean, usually in the examples I give in reading and listening, I will ask fill-in-the-blank or true-false questions. These questions always have a point value to them. So I try to continue with that value (Semi-structured Interview III).

Interpreting and using the assessment results are other significant skills for LAL. Since these were procedures that depended on each other, the data related to them was analyzed together and categorized. The analysis revealed that the junior teacher candidates had strong literacy skills in interpreting and using the assessment results, as in the following table.

Table 22

Interpreting and Using the Assessment Results

STRONG LITERACY
• Discovering the weaknesses (P1-1, P1-2, P2-1, P2-2, P2-3, P2-4, P2-5) • Interpretation about the success of teaching (P1-1, P1-2, P2-1, P2-2, P2-3) • Discovering the strengths of students (P1-2, P2-3, P2-4, P2-5) • Giving feedback to students (P1-3, P1-4, P2-3) • Allowing the students to notice their weaknesses (P1-1) • Finding out whether the students reached the target level or not (P2-1) • Making an interpretation: Incompetency in listening ability (If 2/3 of the class score below 70) (P2-1) • Making an interpretation: Teaching students new content if 75% of students are successful (P2-1) • Comparing each student's total scores with others' (P2-2) • Making a comparison among the classes (P2-2) • Making an interpretation: appropriateness of audio to the students' level (P2-2) • Making an interpretation: the adequacy of students' language exposure (P2-2) • Making an interpretation: the adequacy of practice on the topic of daily routines (P2-2) • Making an interpretation: balance between experience and important grammar teaching (P2-2) • Making interpretations about students' ability to use grammatical structures (P2-2) • Making interpretations about reasons for pronunciation errors (P2-2) • Making interpretations about the reasons for the lack of vocabulary (P2-2) • Making interpretations about problems with fluency and purposeful communication (P2-2) • Making decisions about the parts to be revised, extra questions, activities, and projects (P2-3) • Making an interpretation about learning processes (P2-3) • Making an interpretation about competency in speaking if students do not make important errors (P2-3) • Providing extra help if students continue to make the same errors (P2-3) • Analyzing the results for both each student and the whole class (P2-4) • Questioning the effectiveness of assessment methods (P2-4)

The junior PSETs emphasized that they would make interpretations about the assessment results to discover their students' weaknesses, and then they would try to improve them by drawing on these interpretations. P1-1, P1-2, and P2-1 noted they would do this to enhance the listening skills of young learners. Regarding this, P1-1 will pay attention to "what can be improved" and "where there are weaknesses". Then he will "work on these weaknesses and try to improve them by having the students do listening activities". In a similar vein, P1-2

wrote: “When I analyze the results, I identify exactly where my students are weak. After this, I will modify my post-assessment instruction to focus on my students' weaknesses”. P2-1 also noted that he would prepare a graphic from the most common errors to the least common ones and determine the content that the students would need in the future teaching process for listening. In addition, he added that less than two-thirds of the class scored above 70 points, it would mean that the overall listening skills of the class were inadequate.

P2-2 pointed out that he would follow an inductive procedure from the specific to the general. As the first step, he will conduct an analysis, including the comparison among a students' scores regarding each criterion. Then he will calculate the mean for each component and make interpretations about the weaknesses of the learners according to the predetermined cut-score. By drawing on this analysis, he will also draw conclusions about the effectiveness of the teaching method, as written in the quotation below.

I would do it like this: In a sort of inductive way. First, I would compare the scores of each individual section on each test criterion with the scores of all students. For example, section A has a total of 15 points. How many 15 points out of 30 students? I would calculate the average. Once I had done this for all the sections, I would understand why something was missing from the exam criteria, especially the part that was below a certain criterion. And that leads me to ask: Which element is missing in the lesson, i.e., whether production is missing, whether vocabulary and pronunciation are missing, whether grammar structures have not been learned? I would first compare the parts of the exam in itself. Then, accordingly, I would compare the overall score of the exam among all students. In general, which part was missing? Then, if possible, I would go to a comparison between the classes, get their rates, and find out which class has the highest rate. This allows me to collect data from the classes with the lowest and highest rates. [...] Maybe the children's speaking skills are very good; their coherence of meaning is good, but maybe all of them have weak pronunciation. So I didn't set a good example. Like this, for each criterion, there are aspects of the criteria that will provide me with data, both good and bad. There could be mechanical activities. We said word matching and synonym selection; what's missing in which one? They all constituted an example for me at that point. In the ones that were done well, for example, in writing, the child used the grammar structure well. Most of them made no or very few grammatical mistakes compared to the class. This allows me to interpret that I have taught language structure features well, but that the method I used to teach them was effective (Semi-structured Interview III).

In the semi-structured interview, P2-3 stated that she could detect the students' major weaknesses or the points that she needed to repeat. According to these detections, she thinks that extra questions and activities can be deployed. Moreover, she adds that if there is a subject that she does not realize the students have problems with, she will recover that subject and give extra exercises, tasks, or projects.

P2-4 will analyze the results both individually for each student and collectively in order to see the general level and progress of the class. She will identify the weaknesses that need to be improved and take care to eliminate them in her next lessons. Additionally, she will

question the adequacy of her assessment methods and try to understand them according to the results.

Emphasizing that individual progress is important, P2-5 wrote that he would also take the weaknesses into consideration. To overcome them, he will “write the right prescriptions for the weaknesses”. Although he did not express what these prescriptions would be, he noted that “full efficiency could be obtained from the evaluations”.

Two teacher candidates can also make an interpretation that highlights the learners' strengths in addition to their weaknesses. According to P1-2, the scores of his students reveal how high-quality education he provides and how successful his students are. Furthermore, he notes that these interpretations can also help him improve the way he teaches, and it will make it easier for him to change and innovate by constantly looking at how successful and unsuccessful he is.

P2-4 will also interpret the assessment results to discover the points where the learners are more proficient.

If I think that the student is good at certain things, for example, and then I see a decline or an increase in certain things, I try to evaluate that. I mean, it's not necessarily this thing; it's not why there is a decline. There could also be an increase, and I think this is something to be curious about. Of course, it is a skill that can develop over time, but yes, I think about making such comments like how it happened, and I believe that such things can be very useful (Semi-structured Interview III).

The interpretations will also be related to the quality of teaching and will be used to enhance it. For example, P1-1 will make interpretations about the success of his teaching based on the frequent errors made by his students. Regarding this, he gave the example that if the majority of learners made errors in using appropriate tense or words, it would be interpreted that he could not have provided his students with quality instruction on the grammar subject or the target words. Like P1-1, P1-2 also stated that it was “possible to make a comment on how accurate and successful the teaching method is”. Moreover, he asserted that “this is the best method to evaluate our teaching method”.

P2-1 had the opinion that the assessment content should be compatible with the content covered in the lesson. Although he does not have enough knowledge about how to analyze the assessment results, he states that he will have direct opinions about the effectiveness of the quality of his teaching, and he will decide how much and how well he focused on the assessment content in his lessons.

I'm directly measuring its effectiveness. Is this teaching strategy that we are using suitable for teaching the class in line with the skills I want to measure? I try to train them on grammar, or because I train children on it, not their speech, and so on, but I go and evaluate their speech. When we do this, the results we get in terms of the effectiveness of our teaching strategy, the aspects we need to correct, for example, taking the average of the speaking skill from the speech, or low scores, and so on, depend on what we analyze. As a result, is our teaching strategy right or wrong? Where can it be changed? If we can analyze enough. As I said, I can't think of too many ways to analyze right now. Averages and so on, or the lowest and highest score, in short, it can measure its effectiveness (Semi-structured Interview III).

P2-2 mentioned several times that the assessment results could be interpreted as making decisions about the effectiveness of teaching methods. As expressed by the other teacher candidates, he will also focus on the points where the students have weaknesses, as in the quotation below.

Failing in the grammar category might be an indicator of a student's inability to represent what he/she has learned in performance. Further output practice could be needed. It may also mean I was unsuccessful at teaching the grammar structure itself, so I might need to change my method. If needed, more practice could be added to the lesson that includes the grammar structures that are being taught (Scenarios for Foreign Language Assessment).

According to P2-3, the interpretations based on the assessment results will be beneficial for improving the quality of teaching. He pointed out that detecting strengths and weaknesses would give him the chance "to notice his shortcomings" related to teaching. In order to improve the weaknesses, he will employ some compensation strategies like giving extra activities, making revisions, and assigning projects.

Some of the teacher candidates mentioned that they would give feedback to their students according to their interpretations of the assessment results. The feedback they will provide will focus on both the strengths and weaknesses of the learners, or only their weak points. P1-3, for example, accentuated this in the semi-structured interview and the scenarios, as in the following sentences:

I think it will be better if we give feedback to the students. They can understand their mistakes. For pronunciation mistakes in speaking, you used this word correctly. We can say that you pronounced it. We can say that we should try to use synonyms for words. Don't use the same word. We can report grammatical mistakes. You formed the sentence correctly (Semi-structured Interview III).

P1-4 will also "interpret and inform [the students] in a way that gives them feedback". In the interview, she stated that the students with low scores would be given feedback. For this procedure, she will first identify the frequent errors and create a table of them. Then she will hold a feedback session with the group of learners and focus on these errors. In the scenarios, although she adds that she will not assign scores for the group of young learners' listening skills, she will "point out where they are competent and give feedback on how they can

improve where they are not”. Moreover, she will provide feedback on all four skills for both young learners and adolescents. “Since [her] main goal is to give them effective feedback”, she will “make it clear that they are competent and not competent, and [she will] point out the aspects they need to improve and complete”.

Unlike these participants, P2-3 will pay attention to only the “serious mistakes in terms of both language skills and language components”. According to the interpretations she will make of the assessment results, she will “correct these mistakes in a way that does not make [her] students uneasy”. That is to say, she will provide her feedback in a way that will not stress the learners and demotivate them. Additionally, “if they do not have such serious mistakes, [she] will interpret [her] evaluation positively. If they continue to make mistakes, [she] will provide extra support to make up for [her] students' shortcomings”.

Literacy Related to the Principles Component of LAL

The principles component of LAL involves the teacher’s “awareness of and actions towards critical issues” such as fairness, transparency, democracy, and feedback (Giraldo, 2018). For this component, the findings were categorized into two main sections: awareness of ethical assessment and ensuring ethical assessment. The first of them includes awareness of the features of ethical and unethical assessment processes and the effects of stakeholders such as parents, administrators, and other teachers on fair assessment. The findings, as shown in the table below, reveal that the pre-service teachers in the first case have awareness of these considerations and therefore have a strong literacy in the principles component.

Table 23

Awareness of Ethical Assessment Process

STRONG LITERACY
<i>Unethical Assessment Experience</i> • Unfair scoring (Favoritism) (P1-1, P2-2, P2-3, P2-4) • Assessing what is not taught (P1-1, P1-4, P2-5) • Scoring the exam papers when he was a student (P1-1) • Teachers' not answering the students' questions during the assessment: Inequality (not visiting every class) (P1-1) • Some students cheating because the teacher/proctor leaves the classroom (P1-1) • Not experiencing any unethical assessment (P1-2) • Teacher administering very easy assessments because of unqualified teaching (P2-1) • Teacher helping the favorite student during the assessment (P2-3) <i>Features of the Ethical Assessment Process</i> • Impartiality (P1-1, P2-1, P2-2, P2-3, P2-4, P2-5) • Assessing what is taught (P1-1, P2-4, P2-5) • Prevention of cheating (P1-1, P1-2) • Conducting assessments compatible with the students' level (P2-2, P2-5) • Validity of the assessment process (P1-2) • Adopting non-traditional assessment methods (P1-4) • Assessing the four language skills (Ethical) (P1-4) • No effects of the factors irrelevant to the learning process on the assessment results (P2-1) • Informing the students before the assessment (P2-2) • Getting feedback regarding the assessment (P2-2) • Assessing the content familiar to the students (P2-2) • Concealing the assessment tool before the assessment (P2-3) • Scoring by concealing the students' names (P2-3) • Administering assessments in equal conditions (P2-3) • Teacher being sure that the assessment gives accurate results (P2-4) <i>The Effects of Parents, Other English Teachers, and Administrators on Fairness</i> • Administrators allowing teachers to give exam questions to the students (P1-1, P1-4) • Parents' complaints leading the teachers to administer easy or less comprehensive assessments (P1-1, P2-2) • Parents' pressure on the teacher to raise the students' scores (P2-1, P2-5) • Another teacher's causing the teacher to have negative feelings about the student (P2-3, P2-4) • Teachers not teaching equally (P1-2) • Parents' pressure on the teacher or threats to the teacher (P2-3) • Demands of the administration from the teacher (P2-3) • No effect if assessment is appropriate to aim for and assess what is taught (P2-2) • Other teacher's persuading the teacher not to assess (P2-2) • Preventing teacher from assessing speaking with presentation (P2-4) • No effect of parents in state schools (P2-4)

The members of the first case mentioned some unethical assessment experiences they had in their educational lives. As the issues they disclosed in relation to these experiences had the potential to threaten the fairness of the assessment process, they were categorized as indicators of strong literacy. For example, some teachers' awarding unfair points in favor of favorite students was one of them. Regarding this experience, the teacher trainees pointed out that their teachers gave higher scores to their favorite students, ignored their errors in the exams, and awarded a student a full score (100) without assessing her.

Moreover, some members of the first case reported the teacher administering assessments on subjects s/he had not taught to students as an unethical assessment experience. P1-1 explained that the exams he took involved subjects that had not been covered in the lesson.

In grade 11, I mentioned that all four basic skills are assessed. We had not done any work on speaking and listening until then. This leads us to an unethical type of assessment because we have been reading and writing all our lives, and even writing was very rare in those two years. Asking questions about what was not covered was something that happened in all English classes or all language classes. For example, let me put it this way: books have themes, like business. For example, there are chapters from one to four. For example, we always used to say 1 and 3. There was a question from the 4th unit or a word or phrase that was not mentioned in the book in that theme (Semi-structured Interview II).

In addition to these, some other experiences were also mentioned. As an example, P1-1 expressed that he rated other students' exam papers because his teacher asked him to do that. Moreover, he stated that the teachers not visiting some classrooms during the exam caused some ethical problems since it was advantageous to some students who could ask their questions related to the questions in the exam while some others could not have done so. Some students cheating when the teacher left the classroom and the teacher helping some favorite students during the exam were the unethical assessment experiences P1-1 and P2-3 accentuated.

The features that an assessment should have in order to be called ethical constitute another aspect of the awareness of ethical assessment. With regard to this, the teacher trainees referred to such features as impartiality, assessment of the content taught, prevention of cheating, confidentiality of the students' names and assessment tools, and assessment of language skills through various methods. To exemplify, P1-2 commented that "cheating, if it happens, is a very big problem because I think it will eliminate justice, completely eliminate it". Therefore, he stressed that cheating should be prevented to ensure ethical assessment while noting that he did not have any other idea about how to do it apart from proctoring.

According to P2-2 and P2-5, an ethical assessment is in line with the students' language level. To exemplify, P2-2 wrote, "My evaluation should not exceed my students' comprehension abilities. I have to deliver an evaluation that fits their skills, knowledge, and comprehension". Similarly, P2-5 noted "I approach my students' learning according to their learning and their own level and make my assessments accordingly".

In the Opinion Journal II, P1-4 wrote that "the teacher should look at the situation from a bigger perspective" rather than only the exams. When asked to elaborate in the subsequent semi-structured interview, she gave the explanation below:

We want to make an ethical assessment. I meant that the teacher should not base this ethical assessment solely on the exam. Because the exam is a way of evaluating the student with the paper in front of him/her in a single one-hour period of time. But I think that's not very fair. It's really the classical method. This is a traditional thing. But I think that an ethical evaluation will be made in this way, taking into account the student's participation in the course or the effort and perseverance they have shown in different activities up to that time. [...] In 8 months, these are the weaknesses; these are the strengths (Semi-structured Interview II).

Regarding the ethical assessment, transparency and confidentiality are two crucial issues that the teacher trainees raised. For transparency, P2-2 referred to informing the students about the assessment method and the rubric in advance. P2-3, on the other hand, emphasized that the content of the assessment tool should be concealed before the assessment in order to obtain valid results regarding the students' knowledge or performance. Moreover, she stated that in order to ensure confidentiality, the teacher should not have any information about the identity of the student, and therefore, rating should be done by hiding the names of the students on the exam papers.

Awareness of the stakeholders' possible influence on fair or impartial assessment is the third consideration for the principles component of LAL. The findings in this section focused on how certain behaviors of parents, administrators, or other teachers can threaten the ethical assessment process. Administrators asking the teachers to share the questions with the students in the 8th and 12th grades before the exam is one of these threats to the ethics of assessment P1-1 and P1-4 mentioned. As the reason for this, they stated that the administrators want these students to prepare for the high-stakes exams they will take at the end of 8th and 12th grades.

P1-1 and P2-2 referred to the parents' complaints about the relatively difficult or comprehensive assessments and pointed out that they caused the teachers to conduct easier or less comprehensive ones. For example, P2-2 expressed that the parents care more about

the scores than their children learning the language, and as a result, they dictate to the teacher to prepare and administer less varied assessment tools or to reduce the difficulty of the assessment.

What was the parent's cultural perspective and need for language until that day? Did s/he have a vocational need? These are really huge factors that can have an impact on the teacher. Generally, in our country, such language and other courses are based on that 2-digit number. Score. So it's a passing tool. S/he pays attention to her child's passing the class or learning. Generally, the child's learning turns back to the score. That's why such pressure is mostly placed on the parents. When the question is directed more towards what you lack than what my child lacks, the teacher may be too pressured in this respect and may reduce or weaken the measurement tools. After all, it is that 2-digit score that matters to him. [...] But when, as is often the case in our society, language, especially in the early years, is seen as a passing grade or only a grade in one lesson, the teacher may turn to unethical tendencies in terms of ethics, in terms of his/her own teaching ethics, as he/she will only see it as a numerical element. They empty the content, make the content easier, or ask again for what they have already given because unfortunately, I have seen examples of these (Focus Group Interview II).

As another unethical impact on assessment stemming from the parents, P2-1 and P2-5 draw attention to the parents' pressure on the teacher to increase the students' scores. To exemplify, P2-1 said that he attended four different schools until the upper secondary school and that his parents had a great influence on all of them. He elaborated that they either threaten the teacher or insist that s/he increase the scores "because they all want their children to pass the course even if they fail".

P2-3 and P2-4 accentuated that other teachers may also have a negative impact on the ethics of assessment. To be more precise, they explained that some teachers may create negative impressions about some students in their colleagues' minds, and these may cause the teachers to treat these students impartially and give unfair scores to their performances or exam papers. Similarly, P2-2 stated that other teachers persuade their colleagues not to assess their language skills. According to P2-4, even if the teachers assess their skills, some of their colleagues force them not to use some assessment methods, like making a presentation for speaking assessment.

Ensuring ethical assessment is the second aspect of the principles component of LAL. This includes how pre-service teachers will ensure ethics in language assessment. In this vein, solving the unethical assessment experience, the ethical preparation, administration, scoring, and interpretation of assessment, ensuring fairness according to some factors such as gender, ethnicity, cultural knowledge, and socio-economic level, ensuring democracy and transparency, ensuring fairness in online education, and sharing the assessment results with students, parents, and administrators, ensuring fairness for the students with low vision, and

ensuring fairness in the classrooms consisting of students from different nations are the considerations that are addressed in ensuring the ethical assessment process. Regarding these, the indicators of weak and strong literacy are presented in the table below.



Table 24

Ensuring Ethical Assessment

WEAK LITERACY	STRONG LITERACY
<i>Solution for the Unethical Assessment Experience</i>	<i>Solution for the Unethical Assessment Experience</i>
• Division of grading into two sections (50+50) (P1-4)	• Paying attention to partiality (P2-3, P2-4)
<i>Ethical Assessment Process: Preparation, Administration, Scoring, and Interpretation</i>	<i>Ethical Assessment Process: Preparation, Administration, Scoring, and Interpretation</i>
• Having a scoring table in mind for ensuring equal scoring (pen and paper exams) (P1-1) • Having a dilemma about invigilation (P1-4) ◦ Positive side of invigilation: Preventing cheating (P1-4) ◦ Negative side of invigilation: Demotivating students (P1-4) • Helping the students during the exam (P1-4) • Both ethical and unethical: Concealing the students' names (P1-4) • No effect of scoring the answers (to the same questions) in sequence (P2-1) • Focusing on main skills rather than language components in scoring (P2-2) • Not concealing the students' names (P2-4) • Not comparing the student's score with the mean or a cut point (P2-4) • Assessing the student on an assessment criterion with more points for the skills that s/he is good at (P2-4)	• Administering the same assessment to all students (P2-3, P2-4) • Equalizing the questions among teachers and informing the teachers about the content to be assessed (As a group leader) (P1-1) • Checking the exam paper in terms of the content to be assessed (Not as a group leader or exam producer) (P1-1) • Informing equally all the students about the assessment (P1-1) • Teacher's reading the exam papers without getting help from any student (P1-1) • Paying attention to the student's success rather than personal relationships (P1-3) • Engaging the students more in the lesson (P2-1) • Assessing every skill according to its criteria (P2-2) • Being aware of the students' feelings about unfairness (P2-3) • Allowing the student to show what s/he has learned in the way s/he wants (P2-5) • Giving equal responsibility to the teacher and the student (P2-5)
<i>Ensuring Fairness in Terms of Some Factors</i>	<i>Ethical Assessment Process: Preparation, Administration, Scoring, and Interpretation</i>
<i>Cultural Background</i>	<i>Preparation</i>
• No effect of cultural background on unethical assessment (P2-1)	• Paying attention to the topics used in assessment (P1-4, P2-2, P2-4, P2-5)
<i>Socio-economic Level</i>	• Assessing what was taught (P1-2, P2-4)
• Providing alternative methods for assessment (P2-3)	• Choosing equal roles to be played for role-play activities (P1-1)
<i>Ensuring Democracy in Language Assessment</i>	• Preparing parallel forms of the test (A-B-C forms) (P1-1)
• Allowing the students to answer questions they choose in the exam (P2-2, P2-3, P2-4, P2-5) • Not much possible with young learners (P1-1, P1-2, P1-4) • Preparing parallel forms consisting of questions of equal difficulty (P2-3)	• Objectivity: Lowering the proportion of comments (P1-2) • Appealing to the interests of the students (P1-2) • Paying attention to what the students learned before the assessment (P1-4)
<i>Sharing the Assessment Results with Stakeholders</i>	

Sharing the Results with Parents

- Not the teacher's responsibility (P1-2)
- Sometimes, informing the parents if the students are young learners (P2-1)
- Sharing the results with parents individually when necessary (P2-3)
- Talking to parents if the students have unexpected results (P2-4)

Sharing the Results with Administrators

- Not the teacher's responsibility (P1-2)
- Not informing the administration (P2-1)
- Sharing the results if there is a special reason for doing that (P2-3)
- No idea about whether they should be shared (P2-4)
- Administrators not paying attention to the assessment results (P2-5)

What to Do to Ensure Fairness in Online Education

- Impossible to administer ethical assessment (P1-2, P2-2, P2-5)
- No problem for assessing writing and speaking (P1-2, P1-3)
- No problem for assessing listening (P1-3)

- Paying attention to the students' opportunities before giving homework (P1-4)
- Choosing familiar assessment methods (P2-2)
- Avoiding indirect answers in the assessment tool (P2-2)
- Asking the questions at similar difficulty levels (P2-2)
- Keeping the questions, topics, and content confidential (P2-3)
- Not choosing the content that can create negative attitudes (P2-3)
- Using a comprehensible text or audio involving the known words (P2-4)
- Preparing the assessment tools according to the students' needs, qualifications, and learning (P2-5)
- Preparing assessments compatible with the teaching method (P2-5)

Administration

- Preventing cheating (P1-2, P2-2, P2-4)
 - Asking about the topics appealing to the students' interests (speaking assessment) (P1-2)
 - Allocating more time for the students with physical disabilities (P2-1)
 - Administering the assessment to the students with disabilities after the actual assessment (P2-1)
 - Assessing all students under the same conditions (P2-1)
 - The teacher's standing at a place where s/he can see all the students (proctoring) (P2-1)
 - U-shape seating arrangement or other arrangement styles when there are problems with seeing or hearing (P2-1)
 - Grouping the students according to their levels and personal relationships (P2-1)
 - Asking the questions at similar difficulty levels (P2-2)
 - Avoiding providing any external support during the assessment (P2-2)
 - Treating every student equally during the assessment (P2-3)
 - Giving equal time to the students during the speaking assessment (P2-3)
 - Giving equal tasks to the students during the speaking assessment (P2-3)
 - Being careful not to distract the students while assessing (P2-4)
 - Showing the student that the teacher is listening carefully (P2-4)
 - Treating every student equally and being fair (P2-5)
 - Not causing psychological problems for the students (P2-5)
-

Scoring

- Scoring according to some specific criteria (P1-1, P1-2, P1-4, P2-2)
- Teacher's not including own ideas about students in scoring (P2-1, P2-2, P2-3, P2-4)
- Concealing the students' names (P2-1, P2-5)
- Objectivity (Not tolerating any students) (P2-4, P2-5)
- Scoring by considering what each student did (For group homework) (P1-1)
- Scoring only what the student wrote on the exam paper (P2-3)
- Not expecting extra answers (Scoring according to the intended ones) (P2-4)
- Scoring the *n*th question for every student (P2-4)

Interpretation

- Comparing the students' scores with the cut point (P1-1, P1-2, P2-3, P2-5)
- Monitoring individual development (P2-2, P2-4)
- Considering how much the student learned and used the subject taught (P1-1)
- Unethical: Comparing the student's score with the average (P1-4)
- Comparing the student's performance with the criteria (P2-2)

Ensuring Fairness in Terms of Some Factors

Gender

- Choosing topics appealing to the interests of both genders (P1-1, P1-2, P2-1, P2-2, P2-3, P2-5)
- Unfairness: Some topics appealing to a type of gender (P1-4)
- Using alternative assessment methods (P2-2)
- Using a text not appealing to a gender (P2-4)

Ethnicity/Race

- Choosing more universal topics (not appealing to a specific ethnicity) (P1-1, P2-1, P2-3, P2-4, P2-5)
- Complying with the criteria (P1-2)

General Knowledge

- Choosing a topic that everybody can speak about (e.g., the most exciting moment in your life) (P1-1, P1-2, P2-5)
-

-
- Allowing the students to search about the topic of assessment (P1-2, P2-3)
 - Providing a variety of topics (e.g., 5 topics) (P1-2, P2-4)
 - Not asking about what was not taught (P2-3)
 - Teaching about the topic of assessment (P2-3)

Socio-economic Level

- Assessing the students through cheap materials that every student can buy or access (P1-1)
- Unfairness: Experiential differences (P1-4)
- Unfairness: Economic differences among the students (P1-4)
- Familiarizing the students with the assessment content (P2-2)
- Teacher's providing the students with opportunities for creating equality (P2-3)
- Avoiding the assessment methods requiring expenditure (P2-3)
- Not asking about a book required to be bought (P2-4)
- Administering an assessment without costing the students (P2-4)
- Asking about what was taught in the classroom (P2-5)

Ensuring Democracy in Language Assessment

- Giving students a questionnaire (upper secondary school and tertiary level) (P1-1, P1-2)
- Getting the students' ideas about the assessment methods (P1-3)

Ensuring Transparency in Language Assessment

- Informing the students about the assessment requirements in advance (P1-2, P1-4)
 - Explaining enough of what is expected of the student to do (P1-1)
 - Showing the students their answers and the answer key and explaining their scores (P1-1)
 - Asking questions and conveying clearly what they require the students to do (P1-1)
 - Showing the students their exam papers (P1-4)
 - Recording the speaking performance of the student (P2-1)
 - Teacher's providing feedback on errors and the score (P2-1)
 - Teacher's sharing comments on the student's performance (P2-1)
 - Getting feedback from students about their peers' performance (P2-2)
 - Providing clear and concise instructions (P2-3)
-

-
- Providing criteria being clear and containing exactly what is required (P2-4)
 - Giving the students a table of the criteria (P2-5)
 - Explaining the reasons for and the aim of the assessment (P2-5)

Sharing the Assessment Results with Stakeholders

Sharing the Results with Students

- Giving feedback to the students politely (P1-1, P2-1, P2-5)
- Sharing the results with students individually (P1-2, P2-2)
- Giving the exam papers to the students after scoring (P2-2, P2-3)
- Announcing the scores in the classroom without negatively affecting the students' emotions (P1-1)
- Giving individual feedback to the students after the assessment (P1-1)
- Informing only the students about the assessment results (P1-2)
- Discussing the common errors with the students (P2-2)
- Talking to the students having challenges (P2-4)

Sharing the Results with Parents

- Sharing the results in a meeting or sending messages (P1-1)
- Having individual conversations with parents (P2-2)
- Sharing the results by emphasizing that language use is more important than the scores (P2-5)

Sharing the Results with Administrators

- Obligatory (P1-1)
- Informing the administration about the students' general performance level (P2-2)

Ensuring Fairness for the Students with Low Vision

- Teacher reading the text and questions aloud (P1-1, P1-2, P2-1, P2-2)
 - Using the Braille alphabet (P1-1, P2-2, P2-3)
 - No problem for the speaking assessment (P1-1, P2-1)
 - No problem for the listening assessment (P1-1)
 - Teacher helping the student to write (P1-2)
 - Giving the student more time for assessment (P1-3)
 - Focusing on skills used in oral and written activities (P1-4)
 - Teacher telling the student something s/he watched and asking the student to comment (P1-4)
 - Paying attention to the skills the students are better at (In general) (P1-4)
-

-
- Having the student take the exam under the same conditions as others (P2-1)
 - Helping the student while checking what s/he wrote (P2-1)
 - Preparing an exam paper written in large font size (P2-2, P2-3)
 - Using screen reader software (transferring exam to computer) (P2-2)
 - Preparing a simply designed exam paper (P2-3)
 - Written in contrasting colors (P2-3)
 - Giving the student extra time when necessary (P2-3)
 - Creating an environment where s/he can express herself better (P2-3)
 - Giving the instructions orally when necessary (P2-3)
 - Not assessing with visual materials (Listening and Speaking assessment) (P2-4)
 - Grammar and semantics: More important (Writing assessment) (P2-4)
 - Punctuation and spelling: Less important (Writing assessment) (P2-4)
 - Student taking the assessment with the help of a teacher (P2-5)

Ensuring Fairness in the Classroom Consisting of Students from Different Nations

- Not using Turkish idioms and proverbs (Using international examples in teaching) (P1-1)
 - Using only English for assessments (P1-2)
 - Asking about the content covered in the course (P1-3)
 - Using methods allowing the students to express themselves clearly (P1-4)
 - Writing or talking about a film or commenting about a story or paragraph (P1-4)
 - Not using topics appealing to a specific ethnic group (P2-1)
 - Prioritizing intelligibility in the assessment of speaking skills (P2-2)
 - Choosing content about different cultures for the reading assessment (P2-3)
 - Writing assignments about special days/customs of different nations (P2-3)
 - Using audio about the city or culture of a student from a different nation (P2-3)
 - Giving feedback to students having difficulties in pronunciation due to L1 (P2-3)
 - Not scoring for pronunciation (P2-4)
 - Assessing listening, reading, and writing according to criteria (P2-4)
-

-
- Using the materials related to the target culture (P2-5)
- What to Do to Ensure Fairness in Online Education*
- Asking the students to turn on their cameras during the assessment (P1-1, P1-2)
 - Video recording (P1-2, P1-4)
 - Not paying attention to participation in the lesson (P2-3, P2-4)
 - Not forcing the students to make a presentation (P2-3, P2-4)
 - Sending the student a link that cannot be sent to anyone (Listening assessment) (P1-1)
 - Administering oral rather than written assessments (P1-2)
 - Synchronous computer-mediated communication (P1-2)
 - Using different topics for assessing speaking (P1-3)
 - Observing the students' performance in small groups on Zoom (P1-4)
 - Giving very clear instructions for the assessment (P2-1)
 - Assessing with assignments (P2-1)
-

The category of weak literacy involves actions that can cause unfair, less democratic, and non-transparent assessments. For solving the unethical assessment experience, P1-4 stated that division of scoring into two main sections, each of which would be scored out of 50 points in total, would be a solution. However, this may not solve the problem of unfairness or impartiality unless the teacher gives scores according to specific criteria in a rubric or checklist.

In relation to the scoring phase of the assessment process, the junior PSETs pointed out some behaviors that they will perform to create an ethical assessment process. However, these are the less objective methods or have the risk of impartiality. For example, P1-1 expressed that he would assign scores according to the criteria in his mind rather than the ones in a written rubric while rating the answers on the exam papers.

Let me also talk about pencil and paper tests: In this assessment process, I would prepare a scoring sheet in my head so that every score would be equal. For example, a classical exam. The number of sentences that can be written for that question is 5. For example, if a certain number of points are given to the whole sentence, I would prepare an evaluation scale in my head according to these. If, for example, there is a grammatical error in the sentence but what is wanted to be explained in the sentence is explained, I would prepare an evaluation scale in my head and score equally (Semi-structured Interview II).

P1-4 has the dilemma of hiding the names of the students for scoring without knowing which exam paper belongs to which student. Regarding this, she pointed out that she witnessed one of her math teachers accepted the wrong answer of a student as correct by saying, “That’s exactly how you understood it, so I’m not going to deduct these 20 points,” while s/he was not awarding any scores for the wrong answer given by another student. Based on this experience, she said, “It would be good for the teacher not to look at it from this point of view.” On the other hand, she also had the idea that it was sometimes more ethical to know whose paper it was. In line with this, she remarked “But sometimes, for example, if s/he is a really good student or if s/he really studies, sometimes looking at it can lead to something more ethical in the evaluation of that exam paper”. However, this can be risky for fairness because the teacher will not have enough robust information about whether the student has that knowledge or not and will score according to his or her personal opinions about the student.

P2-4 also expressed that it was unreasonable to conceal the students’ names while scoring the exam papers. As the reason for this, she said “the rule of being a teacher is to be objective” and “if a teacher can’t be confident and says ‘I’ll cover the names of the students

and not know who they are', then for me, at first, a teacher is minus one". Also, she adds that she will certainly not be affected by the identities of the students and will not have any bias "whether [she] reads the paper of the best student in [her] class or the paper of the worst student in [her] class".

Another method that P2-4 said she would apply in terms of scoring, which was questionable in terms of impartiality, was to give more scores to students for the skills that each student was more successful in. She claimed in the interview that this would be more motivating for the students to learn the language and wrote the following sentences in her third journal:

According to my assessment criteria, the main things I need to assess are speaking, listening, reading, and writing skills. Accordingly, for each criterion, I would prepare four different evaluation criteria, with one of each skill having more points but equal points among them. I would divide my students into groups based on the skills I thought they were good at. And I would assess each student on an assessment criterion with more points for the skills that I thought they were good at. In this way, my students would have been evaluated in areas where they could use their achievements well and on equal terms (Opinion Journal III).

Ensuring democracy is another dimension of the ethical assessment process, and the junior teacher trainees will conduct some methods that will not give the learners enough chance to negotiate with the teacher about the use of multiple assessment methods and tools (Giraldo, 2018). Allowing students to choose and answer a certain number of questions from the given ones is the most frequently cited of them. However, this will not enable the students to express their views on assessment tools or methods and choose among them. In addition, three freshmen PSETs uttered that it would not be much possible to achieve this with the group of young language learners. They explained this by stating that the young language learners did not have enough knowledge and experience in language learning, and their decision-making skills were not sufficiently developed, as in the following excerpt.

I agree with him, as [P1-1] said. With something like voting or a survey, we can take the opinions of students at the university or upper secondary school level and make an evaluation without imposition. But like [P1-1], at the primary and lower secondary school level, I think students, as we know, have not developed what we call decision-making skills. It can be a bit more democratic by conducting such surveys and voting. Of course, I am speaking for the higher levels again (Focus Group Interview II).

Sharing the assessment results with the students and parents is also important for an ethical and transparent assessment process. However, some of the teacher trainees either do not prefer to share assessment results with parents at all, or they prefer to do so only for the parents of young learners or when the learners have unexpected assessment results.

Conducting fair online assessments is another dimension about which some junior PSETs mentioned weak literacy indicators. Of these, three teacher candidates pointed out that ensuring fairness in online assessments was impossible, and each of them gave a different reason for this. For example, P1-2 said that it is difficult to achieve fairness because multiple-choice and fill-in-the-blank questions will be used in the assessment of listening and reading skills, and therefore the probability of cheating will be high. On the other hand, taking a more macro perspective, P2-2 stated that online assessments would be negatively affected by inequality of opportunity in education, which would harm the fairness of the assessment. Based on his experiences in online education during the pandemic, P2-5 emphasized that the teaching and learning styles that teachers and students are used to and the curriculum are not compatible with online education.

To tell you the truth, I honestly don't think that the teacher has the opportunity to make an ethical assessment and evaluation in online education. Why? Because of this: First and foremost, our teaching style, our approach to learning, or simply, in the most clichéd way, our curriculum is not suitable for online education. In the training we usually receive, the teacher has always been the leader. We followed him or her and did what s/he did. Now the already limited courses are reduced to 20 minutes, 30 minutes, or 35 minutes with online education. Now the number of students, which was already too many, increased even more. Groups were separated or merged, and the teacher was unable to lead. Students were forced to do things on their own. At first, I was very surprised by this in the first semester. [...] Because this is not our upbringing philosophy. [...] Because they asked us to do something that they almost never asked us to do before. Now I sometimes think to myself: If I were at the university, I would ask, How would this teacher give me this exam? I think about it. I can't think of anything (Focus Group Interview II).

In contrast to these problematic practices in terms of ensuring ethical assessment, the pre-service teachers, who made up the first case, presented practices that were suitable for ethical assessment in many dimensions. They proposed some solutions for their unethical assessment experiences. For example, they pointed out that all students should be subject to the same assessment, and they themselves should be more careful about partiality and their students' feelings about it. Moreover, informing all the learners and other English teachers equally about the assessment, concentrating only on the student's performance by ignoring personal feelings and thoughts about him or her, and scoring and evaluating each language skill in the light of the criteria set for it are the other solutions mentioned by the junior PSETs.

The teacher candidates also mentioned the practices appropriate to create an ethical assessment process in terms of the preparation, implementation, scoring, and interpretation phases of the assessment. In other words, these are the practices that will not cause bias or injustice among the learners. Regarding the preparation phase, they noted they would be

careful in choosing the topics or content for assessment. That is to say, they would not be offensive, advantegous, or disadvantegous for some students and would appeal to their common interests and be familiar to them. To exemplify, P2-5 expressed that using topics related to politics or religion should be avoided, as in the excerpt below.

Now sometimes we give them some topics to discuss, and debate, and these topics can sometimes take things to different places. For example, it can be political, religious or about races. From an ethical point of view, first and foremost, such topics, such debates, or related things should not enter the classroom (Semi-structured Interview II).

They will also pay attention to what is taught before the assessment so that teaching and assessment content will be aligned. If this alignment is not achieved, the assessment will not give accurate results because there will be differences among the students in terms of the knowledge or competence not focused on in the lesson. Moreover, the teacher will create an inconsistent profile in the eyes of his or her students and unnecessarily challenge them. Regarding this, P2-4 stated “In order not to contradict myself and not to put my students in unnecessary difficulty, I would prepare an assessment to evaluate what I have taught”.

Additionally, to ensure fairness in the paper-based exams consisting of multiple-choice questions, P1-1 will prepare the parallel forms of the exams in order to reduce the risk of cheating. Also, P2-2 stressed that these types of questions should be prepared carefully by stating that “if there is a deficiency in the measurement tool that can lead to a direct answer, like ‘a leak in the ship’, [...] it directly violates that ethics in the first place”.

Administering assessment tools or methods is another important phase, and the participants in the first case referred to various practices to ensure the ethics and fairness of the assessment. Preventing cheating is the most frequently mentioned one, and proctoring is the method that the teacher candidates will use to do so. According to P1-2, cheating is a factor that completely destroys fairness and poses a very big problem. Therefore, it must be strictly prevented, and proctoring is the only way to achieve this. Furthermore, P2-1 emphasizes that the teacher, as a proctor, should behave in such a way that the students cannot cheat but are also not distracted by his or her presence in the classroom.

At the time of the exam, the teacher can see everyone, if possible. We don't break the ethics, but maybe a student will try to break them. They can see everyone cheating. But it will not make its presence felt too much; for example, if we talk about the general class order, when we talk about the normal row order, not U, etc., the teacher will be at the front left. That is, he would be standing at the front on the glass side. At most, he can stand there and wait quietly. So that somehow the student focuses. He/she can wait at the back of the classroom, in the sense that he/she makes his/her presence felt less. Maybe this is how the surveillance situation can be made more ethical and correct (Semi-structured Interview II).

Providing equal conditions for all students is another important aspect of the administration phase of the assessment. Regarding this, P1-1 stated that he would play the audio on the computer in all classes for listening assessment. Likewise, P2-1 stated that a U-shaped seating arrangement could be preferred in order to be equal for everyone in assessments that require students to make comments after seeing or hearing material. P2-2 referred to the support provided by the teacher during the assessment and wrote “I would simply avoid providing any external support other than making sure the evaluation guidelines are clearly understood”. Similarly, P2-3 described the help provided by some teachers to some students, such as warning about wrong answers to the multiple-choice questions, as unethical and emphasized that she would stay away from such behavior. P2-2 and P2-3 also stated they would give all students equal speaking time and ask questions of equal difficulty in the speaking assessments.

In order to create a more ethically appropriate scoring process, the pre-service teachers stated that they would implement practices such as not including personal opinions about the student in the scoring, scoring according to certain criteria, scoring by covering the names of the students, evaluating the performance of each student in the group separately, and not expecting answers other than the predetermined criteria. Of these, not including the teacher’s opinions about the student in scoring is the most frequently mentioned practice. Related to this, P2-1 said: “In terms of ethics, the way of analysis and interpretation in which I evaluate the student as a student without including situations such as interest, dislike, and annoyance, that is, without adding any individual feelings, would be ethical.” P2-3 also expressed this, both for the written and oral exams. For the written exams, she emphasized it by saying “Definitely for me, what determines the result that I will give to the student will definitely be what is written on that paper”. For the speaking assessments, she said: “Someone replied with a smile, someone replied coldly. But I got that answer. I don’t take into consideration my own personal thoughts or the personal distance and intimacy I see”.

As a way of not letting his opinions about the student interfere with scoring, P2-5 stated that he could cover his students’ names on the exam papers. While talking about this, P2-5 remembered that some of his teachers did so. Moreover, he drew attention to the possibility of the teacher being affected as a scorer and said, “The smallest thing inevitably affects us. In other words, when you think that a student is making fun of us or making a joke when we write something wrong on the blackboard in class, it inevitably affects the scoring”.

Scoring according to specific criteria is another practice that the two members of the first case intend to implement. For example, P1-2 has the view that objective scoring of written and oral products is difficult, and therefore the teacher's personal comments about these products should be reduced while awarding scores. With this in mind, he wrote that "if the scoring of the students to be assessed in terms of language skills is done by dividing them into criteria, I think the interpretation will remain a little more on the outer circle". In a similar vein, based on her current experience at the university, P1-4 stated that grading should be done using a rubric with the following words:

For example, now we are given rubrics. "5 points come from this". I think this is very logical. After all, at least I can see what I didn't get points for and what I got points for. This kind of evaluation seems more logical (Semi-structured Interview II).

Interpretation of the assessment results is the last but another very important stage of the assessment. In order to make ethical interpretations, the teacher trainees pointed out that they would interpret the scores by comparing them with a cut-off point as in the criterion-referenced assessment since this would not allow any student's score to be influenced by the scores of other students in the group. In relation to this, P1-1 said that comparing a student's score with the mean would involve interpreting a student who scored 70 as failing when the class average was 80. However, according to him, this interpretation should be based on how well students meet the criteria they are expected to fulfill. Likewise, P2-5 preferred to make interpretations according to the cut-off point by expressing that "there was a little more individuality or you had a criterion that you could say what was his or her problem that he got low on or why s/he failed".

As different from this, the two teacher candidates, P2-2 and P2-4, mentioned that they would interpret the students' success by considering their progress over time. Stating that this would be more useful for making detailed decisions about the student's success, P2-2 stated: "Personally, I am in favor of comparing each student's score with his or her own score and evaluating it as successful or unsuccessful". Moreover, he added, "The child is speaking very well. S/he understands some things, but maybe s/he can't put them into production in another direction. So I am a supporter that it would be more appropriate to be considered as an individual". Paying attention to how the student could improve his or her language skills over time was also important for the interpretations made by P2-4. While explaining it, she gave an example of what her math teacher did and pointed out that her teacher paid attention to how much the students could increase their number of correct answers in multiple-choice

questions over time. Based on this experience, she noted that it is more ethically appropriate to pay attention to the results obtained with more than one assessment method and to the individual development of the student in the process.

The junior pre-service teachers mentioned practices that can be considered indicators of high literacy to ensure fairness in assessments in terms of gender, ethnicity/race, general knowledge, and socio-economic level. Of these, the most common practice is to avoid topics or content that would favor or disadvantage a group of learners in terms of these factors in the assessments that the participants will administer to their prospective students. For example, they will use topics appealing to both male and female students instead of such topics as cars, football, make-up, or wars, as in the excerpt below.

For example, in lower secondary school, you asked a question about a war. Generally, history is not something that women are very interested in. I apologize to the other women here, but I take it as "usually". When I look at it here, for example, an action or a war attracts men more, or let me take it as a game. Cars, sports, usually soccer. Men can express their opinions on this more easily. Because those people have given time on that subject. There is a labor; I mean, they have added something. I mean, if I draw a random woman from the street or a random woman from the classroom, how many of them know Jordan? How many of them know so-and-so's athlete? It would be a problem. Yes, it would be a bit unfair. And you can solve that by asking topics or making sure that topics are discussed where each person can have an opinion (P1-1, Semi-structured Interview II).

In terms of the ethnicity/race or nationality factor, the participants stated that they would prefer more universal topics or content to avoid bias and not offending some students. As examples for these, P2-5 mentioned the national and religious holidays and expressed that a teacher might ask the learners to write a text or a poem about them. However, he warns that this can cause differences among the learners "because [their] values, ethnic identities, holidays, thoughts, and geography are different".

Socio-economic level is another important factor that should be taken into consideration for the fairness of the assessment process. Administering assessments in which the materials used do not impose a financial burden on students is one of the ways that pre-service teachers can apply. According to P2-4, "one of the things that creates the greatest unfairness is the disregard of socio-economic status". Based on an experience she had at the university, she stated that some teachers forced the students to buy some books by saying that they would ask the exam questions about the content in these resources. Moreover, she added that this would pose a problem for the assessment of reading skills if the teacher said, "Read this book, and you will be administered an exam in the form of a test about it". As a solution for this problem, she proposed that "the assessments should be carried out in a way that does

not cost students anything,” and therefore the available or free materials should be used in the assessments.

Ensuring democracy and transparency in the assessment process also increases the ethical quality of the assessment. For creating a more democratic assessment, the three junior PSETs pointed out that they could collect data from their high school and tertiary-level students with the help of a questionnaire on what kind of assessment methods or tools they prefer. Regarding this, P1-4 emphasized that this can be done provided that the learners choose only the assessment methods rather than the objectives they are required to attain.

It is not right for students to decide which objectives they want to achieve. Because they tend to choose the simplest one. For this reason, teachers should provide students with a choice of methods to activate whichever outcome they want to activate. For speaking, s/he can create a more democratic environment by listing things like watching a TV series and discussing it or doing this assignment and giving your feedback about it in the classroom, whichever activity the majority of the 30 people agree on (Focus Group Interview II).

As for the transparency of the assessment process, what the teacher trainees will do to carry out a transparent assessment can be considered in three stages: before, during, and after the assessment. Regarding the first stage, providing students with detailed information about the assessment, preparing clear questions and instructions, and sharing the criteria and purpose of the assessment with students were the practices that the pre-service teachers stated they would do before the assessment. During the assessment, P2-1 will record his students' performances for later listening to or watching. In addition, giving the students their exam papers and the answer key, explaining the students their scores, and providing feedback on the errors will be done to ensure transparency after the assessment.

Suggestions and Expectations for the Improvement of LAL

This section includes the junior PSETs' expectations for their LAL development and the suggestions regarding the Exam Preparation in English Language Teaching course. The junior teacher trainees have various expectations from the ELT program to improve their LAL levels. Among these, having knowledge about assessment, practicing assessment techniques, and experiencing different assessment methods are the most frequently mentioned expectations of pre-service teachers. Moreover, they also reported that they expect to create their original assessment techniques and tools as P1-2 explained below:

I would even like this: I mean, let's get to such a level that we can design our own assessment. In other words, I can create our own style and form of assessment. When I start practicing the

profession, depending on the level of the class, I would like to put forward the best quality and the best assessment by evaluating all the environmental conditions. I mean, I would like to reach this level (Focus Group Interview III).

Table 25

Suggestions and Expectations for the Improvement of LAL

<i>Expectations for the LAL Development</i> • Having theoretical knowledge (P1-2, P1-3, P1-4) • Doing assessment practices (P1-2, P1-3, P1-4) • Experiencing various assessment methods (P1-2, P1-3) • Creating own assessment methods (P1-2, P1-4) • Getting knowledge about how to assess the students (P2-2) • Getting information about the success of teaching, problems, and solutions (P2-2) • Being able to enlarge the view about assessment (P2-3) • Practicing the assessment methods taught in the course (P2-3) • Getting feedback about the practice of assessment methods (P2-3) <i>Suggestions for the Exam Preparation in English Language Teaching Course</i> • Ensuring fairness in assessment (P1-2, P1-3, P2-4) • Theoretical information and practice (P2-3, P2-4) • The factors affecting the assessment (P2-3, P2-4) • Compatibility between teaching content and assessment content (P1-2) • How to assess students at different levels of education (P1-4) • Theoretical knowledge (%35) and practice (%65) (P2-2) • Introduction of the assessment methods (P2-2) • How to determine the scoring criteria and aim of assessment (P2-2) • Problems to be faced while administering the assessment methods (P2-2) • Preparing a sample assessment tool (P2-2) • How to administer the assessment methods (P2-3) • Which assessment methods to be used for which situations (P2-3) • Alternative and classical methods appropriate to student profile (P2-3) • The criteria to be used for assessing the students (P2-4)

Junior teacher candidates also made various suggestions for the Exam Preparation in English language teaching course. These suggestions included how to ensure fairness in assessment, providing practical as well as theoretical knowledge, and focusing on factors affecting assessment.

Case 2: Third Graders

This case consists of six teacher candidates studying in the third grade in the ELT department of Gazi University. In addition to the compulsory educational sciences courses and elective professional and general knowledge courses, these teacher candidates take the compulsory applied linguistics courses such as "Teaching English Language Skills I and II", "Teaching

Foreign Languages to Young Learners I and II" and "Teaching Language and Literature I and II", and the elective applied linguistics courses such as "Drama in English Language Teaching/Pragmatics and Language Teaching" and "English Textbook Analysis/Material Design in English Language Teaching". Of these courses, "Testing and Evaluation in Education (TEE)", which is an educational sciences course taught in Turkish, is taken in the 6th semester. The English teacher candidates take this course together with the teacher trainees majoring in other departments, such as Turkish and Social Sciences Education or Mathematics and Science Education. According to the "Course Information" in the "Information Package" section on the Gazi University website, the topics of the TEE course involve "the place and importance of measurement and evaluation in education", "basic concepts of measurement and evaluation", "psychometric properties of measurement tools (validity, reliability, usefulness)", "developing success tests", "analysis and interpretation of test scores", and "evaluation and grading" (Gazi University, 2021). As can be understood from the course content, the TEE focuses on the basic considerations of educational assessment and does not involve training on how language skills can be assessed. Moreover, as some of the participants reported, calculating the indexes of validity, reliability, item difficulty, and item discrimination has an important place in the course.

As for the participants in Case 2, they consisted of 4 female and 2 male pre-service teachers aged between 21 and 23. All of these participants graduated from Anatolian upper secondary schools, and all of them studied preparatory classes at the university. The assessments they experienced until university were usually exams consisting of traditional and non-communicative items such as multiple-choice questions, fill-in-the-blanks, and true-false statements. To be more precise, P3-1, who is a female teacher trainee aged 21, wrote in the Opinion Journal I: "When I think in general, I can say that the assessment methods I have encountered throughout my educational life are individual and non-interactive, that is, they are done on paper." Moreover, she noted that "the assessments in the form of exams and quizzes were common" and explained the individual assessments as practices of assessment not carried out with a peer or in groups. P3-1 started learning English in the 4th grade of the primary school. She remembers that they listened to songs and played games in English lessons. Although she does not remember exactly what the assessments were like in primary school, she guesses that they were "in the form of simple short-answer questions".

In the lower secondary school, her English skills were assessed through assignments such as writing a short paragraph and preparing a poster about a city of their choice, as well as the written exams. While doing these assignments, she got peer feedback on grammar errors. After the teacher reviewed them, the assignments were hung on the class board. Apart from the assignments, the exams focused on reading skills and knowledge of grammar and vocabulary. The items used for the reading assessments were true-false statements and multiple-choice questions while fill-in-the-blanks and true-false statements were employed for assessing vocabulary and grammar.

In upper secondary school, she again took the exams, focusing on her reading skills and knowledge of vocabulary and grammar. The types of items in the exams administered in the 9th and 10th grades were multiple-choice questions, fill-in-the-blanks, true-false statements, and open-ended questions. However, after the 10th grade, she preferred to major in English, and therefore the hours of her English lessons in a week increased. In the 11th and 12th grades, she took mini-exams in which such items as multiple-choice questions, true-false statements, fill-in-the-blanks, and rewriting a given active-voice sentence by making it passive were asked. Additionally, mini-exams were administered for testing the reading skills after they read some classical books, such as *The Old Man and the Sea* while the writing skills were assessed through assignments. However, since they were going to take the university exam at the end of 12th grade, their teacher often preferred to assess them with questions that followed the exam format.

Regarding these assessments, P3-1 thought that her English teachers in primary school used adequate assessment methods within the technological means available at that time. However, she also added, "I think that in primary school, the assessment should be done in a simpler way." With this in mind, she noted that she would utilize web-based activities, games, and portfolios to provide variety in her future assessments.

At the secondary school level, she stated that she would have conducted fewer grammar-oriented assessments. She, therefore, noted that she would assess listening and speaking skills even if they formed a small percentage of the assessment. In order to do this, she would have exploited short, simple, and enjoyable assessment techniques. Additionally, she emphasized the need for a more positive approach, especially in the assessment of speaking

skills, as some students may have difficulties with speaking skills because they are introverts or shy.

Recalling that the English teacher's listening activities in high school did not make sense because listening skills would not be assessed in the high-stakes exam they would take at the end of upper secondary school, P1-3 said that it would be more useful to include listening and speaking skills in the assessment. The reason for this, she said, was that when she started studying in the ELT department at the university, she realized she had challenges with speaking. However, she did not notice this in upper secondary school because her speaking skills were not assessed. In this vein, she wrote "If I were the teacher in high school, I would assess all four basic skills". Within the scope of these assessments, which would have been both formal and informal, she would have made in-class observations and taken short notes, administered mini quizzes on vocabulary and grammar and exams on reading comprehension, asked the students to write a summary or a short paragraph about a reading text or a chosen topic and to give a 10- or 15-minute speech, and given feedback to her students.

P3-2 is a 23-year-old male pre-service teacher. In his opinion journal, he wrote "I think it is vital to mention that I was not assessed in speaking and listening skills until university". At the primary and lower secondary school levels, the assessments were done with true-false statements and open-ended and multiple-choice questions. However, he emphasizes that the questions he called open-ended did not ask the learners to comprehend the text but were "the questions for which the answer was explicitly provided in the text". Furthermore, he was administered the same types of items until the 11th grade, when he became a language major. In the 11th and 12th grades, the assessments consisted of the mock exams, which were prepared in the format of the high-stakes exam to be taken at the end of the 12th grade, and the written exams, which included true-false statements and open-ended and comprehension-checking questions. In addition, the students' participation in the lesson was also taken into account as part of the assessment.

P3-2, who had these assessment experiences, stated that if he were in his English teachers' shoes, speaking and listening skills would have been included in teaching and assessment from primary school to university. As a common practice for all levels of students, he also wrote that he would have preferred to give feedback to students after each assessment

method was used. Moreover, he reported that “I would completely remove open-ended questions from my assessment system, which I see as only measuring knowledge; my aim is not to memorize but to develop my students' interpretation and inference skills”. In the semi-structured interview, when asked how open-ended questions directly measure knowledge, he said that such questions can actually be used to make inferences, but the "open-ended questions" in the exams he has taken so far require the information given in the text to be written exactly in the space below. Therefore, he pointed out that he would have adapted the open-ended questions and true-false statements in a way that would enable students to make inferences about a text they had read or listened to.

P3-3 is a 22-year-old female student teacher. In primary and lower secondary school, she took written exams consisting of fill-in-the-blanks and matching items. “As a result of these exams held two or three times a year, [their] English grades were determined”. In upper secondary school, the exams usually consisted of open-ended and multiple-choice questions. Of these, the open-ended questions were asked in two ways. The first were the questions about writing recipes or giving directions, and they were among the ones she enjoyed answering the most. According to P3-3, “it was an ideal form of assessment as it gave [them] the opportunity to write down what [they] had learned based on the pictures”. The second type were questions about the given paragraphs. She thinks that “these questions are [also] actually a good assessment method”. However, she remembers that her teacher accepted her answers as wrong and deduced her scores because she did not write the same sentences in the paragraphs. Regarding this, she has the opinion that “some teachers made mistakes in the assessment and accustomed their students to readiness”.

Regarding the fact that the assessments in lower secondary school consisted of written exams with fill-in-the-blank and matching items, P3-3 would prefer to use these assessments at the end of each unit to reinforce the learning of the subjects rather than 2 or 3 times a year. As the main reason for this, she explained that she would have aimed "to prevent these exams from creating exam pressure in students". As for the upper secondary school, she noted that multiple-choice and open-ended questions were included in the exams. Considering these types of questions, she wrote that “since I do not think that multiple-choice questions measure English proficiency, I would not include such questions in the evaluation process”. Regarding the open-ended questions, she emphasized that these questions were useful for having a qualified assessment. However, she stated that instead of the teacher asking for the

exact sentences in the reading text, she would also have accepted alternative answers that were related to the question and included the appropriate explanation.

P3-4 is another female participant, and her age is 21. Like other teacher candidates, she started learning English in the 4th grade, and reading, vocabulary, and grammar were focused on during her English learning process until tertiary education. Also, despite the practices sometimes used to improve writing skills in English lessons, she thinks that she was not able to improve them enough. In this context, as compatible with teaching, the assessments she experienced were based on testing reading skills and vocabulary, and grammar knowledge. For example, in primary school, she took written exams that included writing the Turkish or English equivalents of the words, filling in the blanks with the correct forms of the verbs, sorting the elements of mixed sentences, and multiple-choice and open-ended questions. In the 8th grade, assessments were consistent with the high-stakes exam that the students graduating from primary school had to take to get into high school and consisted of multiple-choice questions.

In upper secondary school, she mostly experienced similar assessment methods as in primary and lower secondary schools. However, the difficulty of the reading passages and the sentences for the fill-in-the-blank items increased, and in some exams she was asked to write the synonyms or antonyms of the given words. Also, she notes that “for the first time in grade 10, [she] wrote a short paragraph of 50-100 words on a topic”. In the 11th and 12th grades, she was a language major and took exams involving multiple-choice and open-ended questions about the reading passages.

In addition to the written exams, she would have observed her students’ performances in the classroom and taken detailed notes because she has the opinion that “exam results and similar performance grades are insufficient to evaluate students”. For the upper secondary school, she stressed that she would have absolutely used activities to improve their speaking skills. In this regard, she would have employed portfolios and asked the learners to make short presentations. Drawing on her experiences in preparatory class, she expressed that she would have asked the students to prepare a portfolio or a presentation about the places they visited, like a museum. Moreover, in the crowded classrooms, she would have conducted assessments in which the learners would have completed the tasks in pairs or groups.

P3-5 is a 22-year-old male teacher trainee. He started learning English in the 4th grade. In primary school, he was assessed with written exams in which fill-in-the-blanks and multiple-choice questions were employed. The evaluation involved the scores from the written exams and the scores given by the teacher according to participation in the lesson and whether the learner did the assignment or not. “The content of the assignments was full of mechanical activities, not much different from the exams”. He also remembers that one of his teachers made them do dictation and asked them to give feedback on their peers’ papers.

In lower secondary school, grammar was more focused on in the lessons. In line with this, the assessments were carried out with the written exams consisting of grammar-oriented multiple-choice questions or fill-in-the-blanks items requiring the learners to write the verbs with correct conjugations in the blanks by drawing on the given parts of the sentences.

In upper secondary school, open-ended questions about a reading text were asked in the exams in addition to the multiple-choice questions and fill-in-the-blank items. After being a language major in the 11th grade, he took exams consisting completely of multiple-choice questions because of the washback effect of the university entrance exam. Regarding the assessment of listening skills, he wrote that he took a listening exam in upper secondary school, and this was done because he became a language major.

P3-5 reported that he would have asked open-ended questions to students at all levels instead of multiple-choice and fill-in-the-blank items. He gave the reason that “asking students to choose an option or a word would damage their thinking ability” and their use of English. Furthermore, he would also have made observations “because each student's language development process is different, and with observations [he] can see their development better”. To conduct these observations, he would have adopted a self-made checklist to record students’ use of vocabulary and grammar. However, he also noted that for the language majors in 11th and 12th grades, he would have taught by considering the content of the university exam, and “[he] would have had to use whatever system was used in the university exam as an assessment system”. Additionally, he explained that no time would be taken for improving the learners’ listening or speaking skills if they were not assessed through the high-stakes exam.

P3-6, who is 22 years old, is a female student teacher. In primary and lower secondary school, the focus was on grammar and vocabulary, so the assessment consisted of written

exams and quizzes focusing on them. In the exams and quizzes, multiple-choice questions, fill-in-the-blanks, and open-ended questions were adopted. Of these, the open-ended questions required the learners to write a short text about a topic mentioned in the question while the multiple-choice questions and fill-in-the-blanks tested their grammar and vocabulary knowledge. In the private course she attended to prepare for the high school entrance exam, the teachers focused on grammar and vocabulary, and her vocabulary knowledge was measured with quizzes every week.

In the 9th grade, only grammar was taught in English lessons, and the assessment was carried out through questions asking the learners to make grammatical sentences by using the given elements and fill-in-the-blanks. Moreover, she had to complete the worksheets involving the grammar-oriented fill-in-the-blank items every week and show them to her teacher to get a class participation score. In the 10th grade, assessments were made with multiple-choice questions and fill-in-the-blank items. In the 11th and 12th grades, since she was a language major, the exams consisted of multiple-choice questions similar to those in the university entrance exam.

Considering that her previous teachers had usually focused on grammar and vocabulary and carried out assessments with written exams, P3-6 wrote that she would have assessed her students' all language skills instead of administering "simple and meaningless written exams that were not suitable for the level" of the learners. While explaining how to make these assessments, she remembered her assessment experiences in the preparatory class and in the ELT department. Accordingly, she reported that she would have done speaking assessments in which the learners would have talked about the given prompt, and the teacher would have observed their performances and taken notes. For writing assessments, she would have given two or three paragraph writing assignments to her students in a semester and evaluated them according to a rubric. As for the reading assessments, they would have involved open-ended and multiple-choice questions about the texts she had given. The open-ended questions would have asked the learners to write what they would do if they faced the problem given in the reading text. For the assessment of listening skills, however, she did not know exactly how to assess her primary or secondary school students if she were in the place of her previous teachers. She pointed out that the reason for this was that she had not experienced listening assessments until the tertiary level, and therefore she did not have clear ideas about appropriate assessment techniques appropriate for the level of the students.

Literacy in the Knowledge Component of LAL

Knowledge about the Key Concepts of Assessment

This section involves the findings from the data collected through the first and third semi-structured interviews. The first interviews were conducted in the first weeks of the semester while the third interviews were conducted in the last weeks. The findings are presented separately for each interview in order to depict the effects of the MEE course on the participants' knowledge of educational and language assessment.

If the findings obtained from the first interview are considered, it is clear that a majority of the participants have some serious misconceptions about key concepts such as test, assessment, measurement, and especially reliability, validity, and washback. On the other hand, some of them also have some accurate knowledge about the concepts of test, assessment, measurement, evaluation, inter-rater reliability, and practicality. As for the findings obtained from the third interview, they reveal that some participants improved their knowledge about some concepts such as test, assessment, evaluation, reliability, and washback while others maintained their previous misconceptions or had new ones at the end of the semester.

Table 26

Knowledge about the Key Concepts of Assessment

Categories	Codes (Semi-structured Interview I)	Codes (Semi-structured Interview III)
WEAK LITERACY	<i>Test</i>	<i>Test</i>
	- Multiple-choice questions (P3-4, P3-6) - Method measuring only the students' knowledge (rather than making interpretations) (P3-2)	Limited response items (P3-2, P3-6)
	<i>Assessment</i>	<i>Assessment</i>
	- Making interpretations from the assessment results (P3-1, P3-4) - Using the scores or assessment results (P3-1) - Measurement (P3-2) - Unable to define (P3-3) - Not longitudinal (P3-6) - A quick assessment through assignments (P3-6)	- Evaluation (P3-4, P3-6) - Assessment tools or tasks used to get assessment results (P3-1)
	<i>Measurement</i>	<i>Measurement</i>
	- Observations of the students' skills (P3-1) - Process-oriented (P3-1) - Unable to define (P3-3)	- The existence of a set of criteria (P3-1) - Summative assessment (P3-1) - Supporting the assessment process with numerical data (P3-3) - Evoking the mean and standard deviation (P3-3) - Following procedures during testing according to specific criteria (P3-4) - Testing (P3-4)
	<i>Evaluation</i>	<i>Evaluation</i>
	- Scoring and grading (P3-1) - A final assessment of the tools that measure assessment (P3-2) - The whole assessment process throughout the whole semester (P3-3)	A product-oriented assessment (e.g., final exam) (P3-3)
	<i>Reliability</i>	<i>Reliability</i>
	- The test measuring what is intended to measure (P3-2, P3-5) - Compatibility between the students' success and the assessment results (P3-1) - Accurate interpretation of the assessment results (P3-1) - The consistency between the assessment tool and the assessment results (P3-2) - Research based on a robust foundation (P3-3) - An assessment giving real (accurate) results (P3-4)	- Unable to define (P3-3, P3-6) - Impartiality during the calculation of scores and interpretation (P3-1) - Fair scoring (P3-1) - The same test being administered to the same group at different times (P3-2) - The same test being administered to different groups (P3-2) - The test results reflecting reality or not (P3-4) - Assessing what is taught in the classroom (P3-4) - The questions in the test being related to various subjects (P3-4)
		<i>Intra-Rater Reliability</i>
		- Personal characteristics of the rater (P3-1) - Impartiality (P3-2) - Unable to define (P3-3) - Validity of the assessment tool (P3-4) - The questions being ordered from easy to difficult (P3-5) - The content of the assessment tool involving what is taught (P3-5) - The mutual trust between teacher and student (P3-6)
		<i>Inter-Rater Reliability</i>
		Unable to define (P3-3, P3-6)

• Giving a definitive answer as to whether the student will be assessed fairly or not (P3-4) • Assessment tool appealing to the right sample (P3-5) • Unable to define (P3-6)	• Compatibility between what two collaborating teachers do (P3-1) • Consistency between the content of the assessment tool and what is taught (P3-2) • Reliability between the assessment tools (P3-4) • The security of the exam environment (P3-5) • Consistency between results from two tests (Multiple-choice and open-ended questions) (P3-5)
<i>Intra-Rater Reliability</i> • Unable to define (P3-1, P3-4, P3-6) • Impartiality (P3-2) • Teacher considering the criteria while scoring (P3-2) • Internal consistency of the study (P3-3) • Trust in the teacher doing the assessment (P3-4) • Rater and student giving reliable answers (P3-5)	<i>Methods to Estimate the Reliability</i> • Not having any knowledge (P3-1, P3-3, P3-4, P3-6)
<i>Inter-Rater Reliability</i> • The exam prepared by more than one teacher (P3-1) • The appropriateness of the assessment method(s) for the construct to be assessed (P3-2) • The effects of external factors on research (P3-3) • Ensuring equity between assessment systems used by different teachers (P3-4) • Different assessment tools yielding similar results (P3-5) • An assessment tool yielding the same results for different groups (P3-5) • Getting similar results from a test administered to a different group (P3-5) • Trust in more than one rater (P3-6)	<i>Validity</i> • Unable to define (P3-3, P3-4, P3-5, P3-6) • The consistency between the scores the student took from two exams (P3-1) • The rater's consistent scoring manner in two assessments (P3-1) • Assessing the target skill with the appropriate method (P3-2)
<i>Methods to Estimate the Reliability</i> • Not having any knowledge (P3-1, P3-2, P3-4, P3-6) • Mentioning the alpha index (P3-3) • Not knowing what the alpha index is (P3-3) • Cronbach's alpha (P3-5)	<i>Types of Validity</i> • Not having any knowledge (P3-1, P3-2, P3-4, P3-5, P3-6)
<i>Validity</i> • Unable to define (P3-2, P3-6) • A research study being reasonable (P3-3) • Appropriateness of the assessment methods used (P3-4) • Generalizability of the results obtained from the sample to the universe (External validity) (P3-5)	<i>Practicality</i> • The assessment method/tool being suitable for assessing the target skill (P3-2, P3-4, P3-5) • Being aware of the interpretation of usable assessment results (P3-1) • The assessment tool giving results about the target construct (P3-5) • The questions being useful for the student's learning (P3-6)
<i>Types of Validity</i> • Not having any knowledge (P3-1, P3-2, P3-4, P3-5, P3-6) • Internal validity (P3-3) • Moderate difficulty of the exam questions (P3-3)	<i>Washback</i> • Unable to define (P3-2, P3-3, P3-6) • The impact of exams on student motivation (P3-1, P3-5)

	<i>Practicality</i> • The measurement method helping the teacher measure what s/he wants to measure (P3-2) • Giving accurate results regarding the construct to be assessed (P3-4) • Multiple-choice questions (Not practical) (P3-4) • Availability of the assessment criteria (P3-4) • The results from an assessment tool guiding the teacher on the steps to improve the weaknesses (P3-5) • Unable to define (P3-6) <i>Washback</i> • Unable to define (P3-2, P3-4, P3-6) • Positive or negative effects of the assessment results on the students' behaviors (P3-1) • Proactive inhibition (P3-3) • Retroactive inhibition (P3-3) • Taking actions to improve the student's weaknesses by drawing on the implications of the assessment results (P3-5)	
STRONG LITERACY	<i>Test</i> • Measurement tool (P3-1, P3-5) <i>Assessment</i> • Assignments (P3-5) • Tests (P3-5) • Observations (P3-5) • Portfolios (P3-5) • Measuring the information given to the student through different methods (P3-5) <i>Measurement</i> • Scoring the assessment results (P3-2, P3-4, P3-6) • Measuring the student's progress in language skills and language components quantitatively (P3-5) <i>Evaluation</i> • Making a judgment about the student's success (P3-4, P3-5) • Deciding the student's grade (P3-5) • Deciding what the teacher needs to change in his/her teaching (P3-5) <i>Reliability</i>	<i>Test</i> • The tool used to measure the students' knowledge (P3-1, P3-3, P3-4) • The tool used to assess the student's language skills (P3-4, P3-5) <i>Assessment</i> • An umbrella term involving testing (P3-2, P3-5) • Involving process-oriented assessment methods (P3-3, P3-5) • Assessment involving more than one assessment method (P3-2) • Allowing the teacher to learn about the effectiveness of teaching (P3-5) • Allowing the teacher to learn about the student's strengths and weaknesses (P3-5) <i>Measurement</i> • Scoring according to criteria (P3-2) • Expressing an attribute with symbols or numbers (e.g., height) (P3-5) • Involving scores (e.g., 80, 5 points for each question) (P3-6) <i>Evaluation</i> • Making a decision about the student's success or failure (P3-1, P3-2, P3-4, P3-5, P3-6) <i>Reliability</i> • Consistency among the assessment results (P3-5) • Assessment tool being free from errors (P3-5) <i>Methods to Estimate the Reliability</i>

• Too many questions (Threatens reliability) (P3-5) • A test involving very difficult, medium-difficult, and easy questions (Good for reliability) (P3-5)	• Test-retest method (P3-2, P3-5) • Cronbach's alpha (P3-5) • Parallel forms (P3-5)
<i>Inter-Rater Reliability</i>	<i>Types of Validity</i>
• Two raters giving similar scores to the same essay or assessment tool (e.g., 80 and 82) (P3-5)	• Construct validity (P3-3) • Content validity (P3-3)
<i>Validity</i>	<i>Practicality</i>
• Accuracy of the interpretations of the assessment results (P3-1)	• Time-saving (e.g., multiple-choice questions vs. alternative assessment) (P3-2) • Economical (Easy to access the materials) (P3-2) • Easy to prepare (P3-3) • Easy to administer (P3-3) • Easy to score (P3-3)
<i>Practicality</i>	<i>Washback</i>
• Economical (P3-1, P3-3) • Time-saving (P3-1) • Less effortful or tiring (P3-1) • Appropriate to use for different purposes (P3-3)	• Positive washback effect (P3-4, P3-5) • Negative washback effect (P3-4, P3-5) • Teaching and learning in consideration of the exam (P3-4)

Weak Literacy of the Key Concepts of Assessment (First Semi-structured Interview)

Some pre-service teachers in Case 2 defined the key terms inaccurately or could not give any definition of them. Regarding the concept of “test”, two participants (P3-4 and P3-6) defined it as multiple-choice questions while P3-2 stated that it was a non-communicative assessment method appropriate only to measure the students’ knowledge rather than how they could use the target language. Based on his assessment experience, he stated:

I think the test, as I encountered it in my student life, is a measurement method that aims to measure only knowledge, and by measuring only knowledge, we think that we are getting everything we give to students in an academic year. But as I said, it only measures knowledge, not interpretation (Semi-structured Interview I).

Assessment is another concept about which the participants had misconceptions. While defining it, they referred to different points or could not make any comments. For example, according to P3-1 and P3-4, assessment is the procedure of interpreting or using the students’ scores or other results to improve the quality of teaching. As different from this, P3-2 said it meant the scoring of the data collected through assessment tools or methods while P3-6 commented that this concept reminded her of assignments and defined it as the assessment done through assignments. Moreover, she added, “It's not long-term. It felt like you make something and present it to the teacher, and the teacher evaluates it”. Unlike these pre-service teachers, P3-3 could not make any comments on this concept.

Two teacher candidates were also unable to give accurate definitions or explanations about the concept of “measurement”. P3-1 accentuated that it is process-oriented and involves observing how students use and develop their language skills. On the other hand, P3-3 could not give any answers about it.

Evaluation was defined by P3-1 as scoring and grading. P3-2 commented that this is a final assessment of the tools used in the assessment process. P3-3 and P3-6 pointed out that evaluation is longitudinal and lasts for the whole semester, as in the following excerpt:

Evaluation, as far as I know. I just thought of something like spreading the assessment over the whole process. It is not based on a single exam but on the student's contribution to all learning activities in that semester and year (P3-3, Semi-structured Interview I).

Reliability, validity, practicality, and washback are other important concepts in educational and language assessment. However, as with the other terms, the pre-service teachers either defined these terms incorrectly or did not make any comments about them. For example, reliability is a term for which no participant could give an accurate definition. Regarding

that, they gave definitions of this concept, such as assessment results reflecting the true knowledge or performance of the student, accurate interpretation of assessment results, consistency between the assessment tool and the assessment results, the test measuring what it is intended to measure, or the assessment instrument addressing the right sample. In addition to these, P3-3 remembered that she heard about this concept in the Research Methods in Education course and commented that it means a research study is based on solid foundations.

Intra- and inter-rater reliability are other concepts that the teacher candidates had problems defining. While three participants could not make any comments about “intra-rater reliability”, the others defined it as the teachers’ impartiality, the teacher’s taking the criteria into account while scoring, and the students’ trust in the teacher doing the assessment. Moreover, P3-3 related this term to the research methods and made this guess: “It may be the internal consistency of the research [...]. That is, it is independent from external factors”. As for inter-rater reliability, all participants tried to define it. However, their definitions or comments were not relevant. That is to say, such definitions as a test being prepared by at least two teachers, compatibility of the assessment method with the construct targeted by the assessment, and equality among different assessments conducted by different teachers in terms of the accuracy of the results and fairness were made. Additionally, P3-3 stated that this concept is the opposite of intra-rater reliability and guessed that it refers to the effect of external factors like gender on the research. P3-5 also made two comments about the term. The first of them was that inter-rater reliability means that different assessment tools yield similar results while the second was that the same assessment tool yields similar results when it is administered to different groups.

The methods for estimating reliability were asked of the teacher candidates, and they were expected to refer to methods such as test-retest, parallel forms, split-half, KR-20, KR-21, and Cronbach’s alpha and explain in which cases they are used. However, four of them stated that they had no knowledge about it while two other participants mentioned one of these methods but could not explain what it was. As an example, P3-5, who reads research articles as extensive reading because he wants to pursue an academic career after graduation, emphasized that he remembered Cronbach’s alpha since he met it in some articles. However, he could not give an explanation about the features specific to this index.

What I remember is that I think some reliability tests are done in SPSS. Was it even Cronbach's alpha? I remember that. I don't like it, to be honest. I don't like to read things with statistical, numerical, or quantitative data. More qualitative stuff. I like the qualitative ones more because there are interviews and so on, but sometimes we have to read them. There is a Cronbach's alpha I remember from there (Semi-structured Interview I).

Validity is another crucial term about which every pre- and in-service teacher should be knowledgeable. However, the findings revealed that the participants did not have sufficient knowledge about this concept. In other words, two pre-service teachers in Case 2 were unable to make any comments while three of them made irrelevant definitions of the concept. Of these, P3-3 and P3-5 remembered the content covered in the Research Methods in Education course. In line with this, P3-3 defined validity as the rationality of a research study while P3-5 considered the concept as the external validity of the research and gave the example of the generalisability of the results obtained from Gazi University to other universities in Ankara or in Turkey. According to P3-4, it means that the assessment methods used are suitable for the assessment of the language skills, and gave the definition below:

Whether it's valid or not. But it's very much the same concept. I think you understand. I mean, is it appropriate or not? You know, we can think about it. [...] To use the same assessment methods in two different independent courses. It would be a bit illogical to use the same methods. For example, the criteria we use to measure language skills or other criteria should not be the same in another course. Because language learning is a process and I think we cannot evaluate it only by measuring knowledge, I think that a measurement method that focuses only on such technical things should not be the same as the evaluation methods we will use in our field. I think it would be invalid (Semi-structured Interview I).

As with the definition of validity, the participants were not familiar with the types of validity or validity evidence. Five of them pointed out that they did not have any knowledge about them. On the other hand, P3-3 stated that she only remembered the concept of internal validity, although she did not know what it was. She also added that the use of strong distractors in multiple-choice questions is a factor affecting internal validity.

Practicality is another concept about which the four members of Case 2 have some misconceptions or have no knowledge at all. While this term meant nothing to P3-6, other participants mentioned different points. To exemplify, P3-2 expressed that this concept means that the measurement method or instrument allows the teacher to measure what he/she wants to measure. According to P3-4, if an assessment method or instrument gives accurate results regarding what the teacher aims to measure and is compatible with the construct targeted by the assessment, it means that method or tool is practical. As an example of this, she commented that the multiple-choice questions are not practical since they are not appropriate to assess how the learners can use the language skills. Furthermore, she made

another explanation and expressed that the availability of detailed criteria before the assessment and the teacher scoring according to these criteria would be an example of practicality. As different from these definitions, P3-5 referred to the relationship between teaching and assessment and pointed out that it will be practical if the assessment results can guide the teacher about further teaching.

The teacher candidates do not know what the concept of washback is. In other words, three prospective teachers (P3-2, P3-4, and P3-6) pointed out they did not have any guesses or comments about this concept. As for the answers given by others, P3-1 commented that this includes the positive or negative effects of assessment results on students' behavior. She added that the students might change their behaviors according to the assessment results or the scores given by the teacher for their participation in the lesson. Based on the experience of online education during the pandemic period, she gave the example that their professors asked them to turn on their cameras and participate in the lesson, but they did not do so and got low scores for class participation, and then they decided to attend the class by turning on their cameras. P3-3 remembered the concepts of “proactive inhibition” and “retroactive inhibition” taught in the educational sciences courses and guessed that the washback effect might be related to these types of inhibition. According to P3-5, washback effect is the teacher using the interpretations about the students' language skills based on the assessment results in order to improve their weaknesses in language skills.

Strong Literacy of the Key Concepts of Assessment (First Semi-structured Interview)

The participants in Case 2 also gave some accurate definitions or explanations of some key terms specific to educational and language assessment. For example, P3-1 stated that the concept of test involves assessment tools such as written and oral exams and assignments. In a similar vein, P3-5 defined a test as “a resource and a tool that we can [use to] measure [...] the knowledge given”.

Assessment was also accurately explained by P3-5 by referring to various techniques of assessment. In this explanation, he emphasized that the students' knowledge can be assessed in the process through portfolios, assignments, tests, and observations, as given below:

Let me put it like this: attempting to measure a given piece of information with many different methods, not only tests but also many different methods. We can also include observations in this. Also, assessment can include portfolios made by the student. Homework may be included in this assessment. Based on the observations made by the teacher in the classroom, participation can also be included (Semi-structured Interview I).

Two pre-service teachers gave accurate definitions about the concept of measurement. Teacher giving scores is the common feature that these participants mentioned while explaining it. To exemplify, P3-2 stated that the procedure of scoring the data collected through a test is measurement. Similarly, P3-6 pointed out that measurement means the quantification of the assessment data, as in the excerpt below:

For example, a test. I can think of a written test. I think there is a certain number. [How? Can you elaborate a bit?] For example, a student got 90 on a written exam. 90. That's just a measurement. Measurement in numbers (Semi-structured Interview I).

Only one participant had accurate knowledge about the concept of evaluation. P3-4 explained what evaluation is by giving an example. She stated that the procedure of making a judgement by comparing the assessment results with the criterion is called evaluation. Also, she gave the example of this judgment: if a student gets a 49 on an exam and the cut-point score is 50 for passing a course, it is judged that the student failed the course.

P3-5 provided correct explanations about the inter-rater reliability. Although P3-5 gave two inaccurate definitions of this concept, he could also explain it successfully in the third one. He commented that the inter-rater reliability means the consistency between the scores assigned by at least two raters to the same essay or the data from an assessment tool. He also exemplified it with these words: "You read an essay, and you gave it 80 points. I read it, and I gave it 82 points. We gave close scores. We had reliability between our evaluations. I think there is reliability here because we gave parallel scores."

Only P3-1 was able to give a correct definition for validity. She referred to the validity of the evaluations made from the measurement results in the definition given below:

We did the measurement. When we interpret these measurement results, we make a judgment, the validity of that judgment, that is, how accurate or valid that judgment is (Semi-structured Interview I).

As for practicality, P3-1 and P3-3 mentioned relevant features to explain it. For example, they stated that an assessment method or tool should be economical. Moreover, P3-1 stressed that the assessment process involving practical methods or tools should be time-saving and less tiring for the assessor. P3-3 also added that practicality involved the assessment materials being appropriate to use for different purposes in different assessments.

Weak Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)

According to the findings from the third semi-structured interview, some of the teacher candidates in Case 2 lack sufficient knowledge about the key concepts related to assessment. Test is one of them, and two PSETs (P3-2 and P3-6) has some inadequate literacy about it. In other words, P3-2 defines test as a paper-based material including fill-in-the-blanks and true-false items. Similarly, P3-6 also had the opinion that test refers to a paper-based instrument. However, she stated that it included multiple-choice questions.

Assessment is a key concept about which three of the participants do not have accurate knowledge. Of these, P3-4 and P3-6 considered the concept of assessment as the same as evaluation and defined it as a situation of making various interpretations and decisions based on measurement results. On the other hand, P3-1 expressed that assessment refers to the tools or tasks deployed for obtaining results regarding the students' knowledge or performance.

Three PSETs could not give correct definitions of measurement and their answers were divergent. According to P3-1, for example, measurement means either the existence of a set of criteria that will be regarded in the assessment or the summative assessment itself. As understood from these definitions, she did not have any exact knowledge about it. P3-3 also made an irrelevant definition of this concept. Stating that measurement is more quantitative and reminds her of mean and standard deviation, P3-3 defined it as "evaluation process being more supported by the numerical data." Like P3-1, P3-4 also was not sure what measurement is. Therefore, she also made two different definitions regarding this concept. The first of these was, "The things we use and the operations we perform during testing should be done according to certain criteria," while the second was that testing procedure means measurement.

P3-3 also lacked of the correct knowledge of what the concept of "evaluation" was. Stating that all the terms asked in the interview looked like similar, she commented that it is a product-oriented process. In line with this, she mentioned final exams as the examples of measurement.

Reliability is another concept that five out of six participants do not have accurate and sufficient knowledge about. To put it more clearly, P3-3 and P3-6 did not have any guesses

regarding this concept. On the other hand, others defined reliability by referring to aspects such as unbiased scoring and interpretation, the administration of a test to various groups, and the relevance of test results to real life.

Furthermore, the participants could not define the intra- and inter-rater reliability correctly. Regarding these concepts, P3-3 could not give any comment. However, the others tried to guess them. For instance, they referred to this concept as a personal characteristic of the rater, the impartiality of the teacher, a feature of the test, or a personal relationship. Likewise, they could not give accurate definitions about the inter-rater reliability. They considered it as a concept referring to the consistency across multiple elements like tests, scores, or contents. As compatible with their weak literacy in defining the reliability, four teacher candidates could not provide any answers regarding the methods of reliability estimation.

Validity was also a challenging concept for the third-year teacher trainees. That is to say, four of them were unable to comment on it. On the other hand, P3-1 defined the validity as a consistency either among the scores or the rater's scoring manner. However, P3-2 gave an example implying that he had some knowledge about the construct validity although he could not give an accurate definition of it.

Five pre-service teachers also did not have any knowledge about the types of validity. In addition, although they had some comments on practicality, these were far from the actual definitions of the concept. Regarding the concept of washback, P3-1 and P3-5 defined it as the negative effects of the exams that they take on their motivation.

Strong Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)

The findings revealed that the PSETs also could give accurate definitions or explanations about some key terms. Test was one of them, and four PSETs could explain it correctly. In their definitions, they mentioned it is an instrument that can be exploited in order to collect information regarding the students' level of knowledge and performance in a specific domain.

Assessment was also correctly defined by three student teachers addressing its different features. For example, they reported that assessment is a broader concept involving testing

and it involves the use of a number of methods or techniques. Moreover, they also referred to its process-oriented nature. P3-5, who is literate enough about this concept, also stated that assessment enables the teacher to identify the strengths and weaknesses of his/her students and to get knowledge about the success of his or her teaching process and methods.

It is something that helps us to find out how much the instruction we give to the student affects the student, how much he/she has learned, what he/she is good at and what he/she is not good at. [...] Not a test, observations, other different things (P3-5, Semi-structured Interview III).

As for measurement, which involves the quantification of the learners' performance or competence, is the concept that three participants know. In their explanations, they mentioned scoring or labelling the performance or product in question. For example, P3-5 provided the following definition:

As far as I remember from the lesson, I know that we look at an observation or characteristic of the students and express it with numbers or symbols. For example, we say that Ahmet's height is 1.60. This is a measurement that we express (Semi-structured Interview III).

Evaluation is the concept that most participants answered correctly. In their definitions, they explained it as the process of deciding whether students have succeeded or failed based on the measurement results. Additionally, P3-5 had the sufficient knowledge about the reliability and defined it as "measurement results are consistent with each other". Moreover, he added that it also refers to the assessment tool being free from the measurement errors. However, he could not give a correct definition of the inter-rater reliability.

P3-2 and P3-5 know the name of the methods to estimate the reliability index. For example, test-retest method was mentioned and explained by both of them. However, P3-5 also referred to Cronbach's alpha and alternate forms. He stated that he remembered these methods from the MEE course and that he also encountered these types of reliability in the articles he read. Regarding the types of validity, only P3-1 could mention the names of the validity types: content and construct validity. However, she could not give any explanation about them.

Practicality is a concept about which P3-2 and P3-3 have sufficient knowledge. In their statements on Practicality, P3-2 and P3-3 state that in order for an assessment method or tool to be considered practical, it should be economical, time-efficient and facilitative at various stages of the assessment. While giving her explanation, P3-3 said that when she thinks of practicality, the first thing that comes to her mind is the teaching materials and she transferred the features of the practical materials to the assessment:

When I think of usefulness, I immediately think of something like this: material. That immediately came to my mind. How is usefulness in measurement? Let me think about that. Actually, they all seem to be similar. [In the material, a material that is useful. You can actually transfer this to measurement]. So if we can apply a material well in the classroom, it is useful. The measurement tool is the same way. If it is not difficult or troublesome to apply. It can be. Then again, it can be a teacher's process of preparing an assessment tool. [Time consuming.] Yes. In terms of economy or. Exactly in terms of time. These things come by associating them with the material. Also, for example, it can be something we mentioned before: Scoring it. For example, the scoring process can also be included here. [If it is easy to score it.] Yes. It is useful. [For example, multiple-choice.] I think it is more useful (Semi-structured Interview III).

P3-4 and P3-5 defined washback correctly. They referred to the negative and positive effects of the examination on learning and teaching. For example, P3-4 stated that the English teachers in Turkey adapt the content of instruction in an exam-oriented way if their students are going to take a high-stakes exam, and that the students' learning has also become more exam-oriented.

Knowledge about the Types of Assessment

The teacher candidates in Case 2 have some misconceptions about types of assessment while they also have some accurate knowledge or conceptions of them. The findings in this section are also presented under the categories of weak and strong literacy, as in the previous one.

Table 27

Knowledge about the Types of Assessment

Categories	Codes (Semi-structured Interview I)	Codes (Semi-structured Interview III)
WEAK LITERACY	<i>Summative Assessment</i>	<i>Summative Assessment</i>
	• A general assessment of the assignments (P3-2, P3-6) • A summary of all the students' scores by calculating score ranges (P3-1) • A summary of the whole year (e.g., the participation in the lesson throughout the whole year) (P3-2) • Unable to define (P3-3) • Process-oriented assessment (P3-4) • Assessment based on students' overall performance (P3-4) • Summary of the assessment results done during the semester (P3-4)	• Making interpretations according to the frequent results (P3-1, P3-6) • Process-oriented (P3-2, P3-5)
	<i>Formative Assessment</i>	<i>Formative Assessment</i>
	• Unable to define (P3-1, P3-3, P3-4, P3-6) • Formal and legal assessment (e.g., midterm and final exams) (P3-2) • Assessment done at the end of the unit or midterm (P3-5)	• Unable to define (P3-1, P3-6) • Assessment methods with definite answers (multiple-choice questions, true-false statements) (P3-2)
	<i>Diagnostic Assessment</i>	<i>Diagnostic Assessment</i>
	• Teacher giving feedback to the student about strengths and weaknesses after the assessment (P3-2) • Detecting the students' frequent errors as a result of the in-class observations (P3-6) • Classroom-based assessment (P3-6)	• Unable to define (P3-3)
	<i>Dynamic Assessment</i>	<i>Performance-based Assessment</i>
	• Unable to define (P3-1, P3-4) • Giving a plus or minus according to the student's participation in the lesson (P3-2) • Involving formative, summative, diagnostic, and performance-based assessment (P3-3) • Assessing everything the student does in the education and training process (P3-5) • Students not being aware of the assessment, a quick assessment (P3-6)	• Scoring only according to the student's performance in assessment (P3-1)
		<i>Dynamic Assessment</i>
		• Unable to define (P3-1, P3-3) • Teacher's observation about the student's participation (P3-2) • Assessment done from the beginning to the end of the semester (P3-4) • Involving the changes made according to the feedback from the students (P3-5) • Assessment done during the activity (P3-6)
		<i>Criterion-referenced Assessment</i>
		• Scores assigned for each subsection (e.g., grammar) (P3-3, P3-6) • Teacher giving an extra score because of the student's participation (P3-2)
		<i>Norm-referenced Assessment</i>
		• Unable to define (P3-1, P3-3, P3-6) • Dichotomous scoring (e.g., multiple-choice questions) (P3-2) • Testing done for detecting the student's general status without a criterion (P3-4)

	<i>Traditional Assessment</i> • Open-ended questions (P3-2, P3-4, P3-6) • Assessing the students from different perspectives (P3-4) • Process-oriented (P3-4) <i>Alternative Assessment</i> • Quizzes (P3-6) <i>Norm-referenced Assessment</i> • Unable to define (P3-1, P3-4, P3-5) • Inference questions (Involving different inferences and answers) (P3-2) • Scoring according to personal judgment (Holistic rubric) (P3-3) • Assessment based on general observation (P3-6)	
STRONG LITERACY	<i>Summative Assessment</i> • Used to specify the level (P3-5) • Used for deciding whether the student can pass or fail (P3-5) • Used for determining what the student has learned after covering a unit (P3-5) <i>Formative Assessment</i> • Teacher asking a question (P3-5) • Quizzes (Example given by the instructor of the MEE course) (P3-5) • Making interpretations about the weaknesses and the teaching methods and making changes to improve them (P3-5) <i>Diagnostic Assessment</i> • Making judgments about students' strengths and weaknesses based on the assessment results (P3-1) • Teacher making a diagnosis regarding the student's learning process (P3-3) • Identifying the students' strengths and weaknesses in detail (P3-4) • Assessment to determine the student's progress in the process (P3-5) <i>Performance-based Assessment</i> • Speaking (P3-3, P3-5, P3-6) • Writing (P3-3, P3-5, P3-6) • Involving all activities related to the lesson (e.g., participation, doing homework) (P3-1) • Assignments (P3-1)	<i>Summative Assessment</i> • Product-oriented assessment (e.g., midterm and final exams) (P3-3, P3-4, P3-5) <i>Formative Assessment</i> • Process-oriented assessment (e.g., project assignments, portfolios, quizzes, groupwork, feedback) (P3-3, P3-4, P3-5) <i>Diagnostic Assessment</i> • Assessment giving detailed information about the student's strengths and weaknesses (P3-1, P3-2, P3-4, P3-5, P3-6) <i>Performance-based Assessment</i> • Assignments (oral and written) (P3-2, P3-3, P3-5, P3-6) • Tasks (P3-2) • Portfolios (P3-4) • Role-play performance (P3-5) • Observation (P3-6) <i>Traditional Assessment</i> • Multiple-choice questions (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6) • True-False statements (P3-1, P3-2, P3-4, P3-6) • Fill-in-the-blanks (P3-1, P3-2, P3-3) • Paper-based (P3-1, P3-2, P3-4) • Used to make clear detections and giving scores (P3-5) <i>Alternative Assessment</i> • Performance tasks (P3-1, P3-2, P3-4) • Project assignment (P3-2, P3-3, P3-6) • Role-play (P3-3, P3-5, P3-6) • Process-oriented (P3-1, P3-5) • Portfolio (P3-5, P3-6) • In-class observation (P3-2)

-
- Portfolios (P3-1)
 - Projects (P3-2)
 - Involving the production of a hand-made material (P3-2)
 - More practical (P3-3)
 - An assessment based on observations made during the semester (P3-4)
 - Assessing the performance of the student throughout the semester (P3-5)
 - Willingness to participate in the activities (P3-5)
 - The performance of the student during the pairwork and group work activities (P3-5)
 - The students' works done during the pairwork and groupwork activities (P3-5)
 - Classroom-based assessment (P3-6)

Traditional Assessment

- Paper-based exam (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6)
- Multiple-choice questions (P3-2, P3-4, P3-5, P3-6)
- Fill-in-the-blanks (P3-2, P3-5)
- Limited response items (P3-4, P3-5)
- Legal or formal assessment (P3-2)
- Yes-No questions (P3-2)
- Exam-based (P3-4)

Alternative Assessment

- Portfolio (P3-1, P3-2, P3-3, P3-4, P3-5)
 - Assignments (P3-5, P3-6)
 - Innovative assessment methods (P3-1)
 - Internet-based assessment methods (P3-1)
 - Focusing on the student's performance (P3-2)
 - Project assignments (P3-2)
 - Oral exams (P3-2)
 - Making a presentation (P3-2)
 - Demonstrations (P3-2)
 - Peer-feedback (P3-2)
 - Speaking exams (P3-3)
 - Essay-style assignments (P3-3)
 - Search activities (P3-3)
 - Preparing a board (P3-4)
 - Role-play (P3-4)
-

- Debate (P3-3)
- Feedback on the students' products (P3-5)

Criterion-referenced Assessment

- Success or failure according to a cut-point score (e.g., 60) (P3-1, P3-5)
- Testing done according to a rubric or checklist (P3-4)

Norm-referenced Assessment

- The student's status among peers according to his or her score (P3-5)
-

-
- Allowing the students to produce by using creativity and expressing themselves (P3-4)
 - Teacher's observation (P3-5)
 - Focusing on the students' participation in the lesson (P3-5)
 - In-class performance (P3-6)

Criterion-referenced Assessment

- Scoring according to criteria (P3-2, P3-3, P3-4, P3-5, P3-6)
 - Whether the students are able to accomplish the criteria (P3-1)
 - Having exact answers or criteria (P3-2)
-

Weak Literacy of the Types of Assessment (First Semi-structured Interview)

The teacher candidates gave definitions and explanations or made comments indicating that they lacked the necessary knowledge about various types of assessment. For example, summative and formative assessment are two types about which pre-service teachers mostly lack accurate knowledge. Of these, the summative assessment was generally defined as an assessment consisting of a comprehensive summary of the students' performances, assignments, or scores from various assessments. To exemplify, P3-2 and P3-6 defined it as the procedure of the teacher collecting all the assignments the learners did throughout the whole semester and giving a score as a result of the examination of all these assignments. As different from this, P3-1 made a guess by drawing on the word "summative" and defining it as an analysis of all students' scores in a class and determining their percentile ranks. In contrast to summative assessment, formative assessment did not mean anything for more student teachers. In other words, four participants could not give any answers to define or explain it. According to P3-2, on the other hand, formative assessment means formal or legal assessments like midterm and final exams, which are administered regularly.

In my opinion, formative assessment takes the form of a formal exam—not an exam, but a formal measurement, measurement, and evaluation. For example, the lines prepared for written exams or presentations—let me call it the process of filling those lines. It is a type of formal evaluation, I should say, so for university students, which is done twice a year by the teacher who makes the observations and at the end of the semester on paper (Semi-structured Interview I).

Diagnostic assessment is the type of assessment about which only P3-6 had misconceptions. She defined it as an assessment done to detect the student's frequent and repeating errors. Regarding that, she gave the example that a student makes the same errors while speaking, and the teacher observes that student and diagnoses these errors. Moreover, she added that this can also be done through written products.

Dynamic assessment is another type of assessment, and all student teachers lack the necessary literacy about it. While two participants were unable to define it, the others provided various answers. For example, P3-2 and P3-6 guessed that it was a spontaneous assessment done during the lesson. In addition, P3-2 stated that the teacher's taking notes in the form of plus and minus according to the students' participation in the lesson is dynamic assessment. According to P3-3, dynamic assessment is more comprehensive than other types of assessment and includes summative, formative, diagnostic, and performance-based assessments. P3-5 said that although he had heard of dynamic assessment before, he had not

read about it and therefore was not sure what it was. Nevertheless, he made a guess by drawing on its name and defined it as the fact that “everything done in the education process is measured dynamically without skipping anything”. That is to say, the student’s performance while doing an activity, his or her participation in the lesson, and how s/he uses his or her language skills are all elements of dynamic assessment.

The NRA is another type of assessment for which the participants are not literate enough. To be more precise, three teacher candidates do not have any knowledge or guesses about it while the others have some misconceptions. Among these prospective teachers, P3-2 remarked that the questions requiring the students to make inferences and not have exact answers constitute the NRA. On the other hand, P3-3 considered the NRA in terms of scoring and claimed that the scoring process in the NRA is based on the teacher’s intuitive judgment rather than specific criteria. Therefore, according to P3-3, the NRA is “the assessment where we never know what the teacher scores”. P3-6 also briefly commented that it is “an assessment based on a general observation” rather than the detailed criteria set to evaluate the student's performance.

Strong Literacy of the Types of Assessment (First Semi-structured Interview)

The participants also have some accurate conceptions or can make correct guesses about the types of assessments. For example, summative assessment, which is conducted at the end of a unit or a semester to determine how much the students have learned, was accurately explained by P3-5. In his statements, he mentioned three purposes for which the summative assessment is carried out. The first one was that the summative assessment was conducted to detect the students’ levels, and he pointed out that he read it in an article. The second purpose he referred to was that it was used to determine whether students passed or failed a course while the last one was to detect what students had learned from a unit after it had been covered.

P3-5 was also the only teacher candidate who was able to give correct answers to define the formative assessment. While making his explanation, he provided the example given by the instructor teaching the MEE course and said the following sentences:

This one too, I guess. I remember the example I saw in the lesson when the teacher asked a sudden question during the lesson. The measurement teacher also gave such an example. Quizzes

in English, for example, were an example of formative assessment, he said. Planned or unplanned. As I said, it is an assessment that can be done either in the form of a planned quiz or an unplanned pop quiz. After that, if it was done at the end of a unit or if a general mid-term assessment was made, as a return of this, the teacher's own, now I should not say teaching style, methods, or feedback received in the lesson. And did you understand this? Or, after asking questions to understand this, if he/she realized that it was not understood, he/she would teach it again. I remember this as this kind of measurement and evaluation (Semi-structured Interview I).

The participants in Case 2 also gave accurate definitions or explanations regarding the performance-based assessment, which allows the assessor to directly observe the assessee's actual or simulated performance in language skills (Brown, 2005, pp. 10-11). In this regard, the assessment of speaking and writing skills was mentioned by P3-3, P3-5, and P3-6 by emphasizing that the performance-based assessment focuses on production through the use of language skills. In a similar vein, P3-1 stated that the assignments and portfolios, which require the learners to speak and write in the target language, will be employed in the performance-based assessments. P3-2 remembered his primary school years and stated that the constructivist approach had just entered his educational life at that time, and he did some project assignments. Based on this experience, he said that the project assignments are a part of performance-based assessment.

Traditional and alternative assessments are the types of assessments about which all the participants have appropriate conceptions. Regarding the first one, all of them reported that traditional assessment consists of written exams. Furthermore, they referred to the types of items used in these exams and emphasized that multiple-choice questions, fill-in-the-blanks, or limited-response questions are deployed for measuring the students' knowledge.

Alternative assessment, which is accepted as more authentic than traditional assessment, is another type that the teacher trainees are able to explain or define accurately. Regarding this type of assessment, the portfolio is the alternative assessment technique that the participants think of the most. In addition, two participants also remarked that the assignments were also adopted in the alternative assessments. Role-play, demonstration, presentation, observation, and lesson participation are also other elements cited by the participants while explaining this assessment.

Weak Literacy of the Types of Assessment (Third Semi-structured Interview)

The teacher candidates in Case 2 had some misconceptions about or no knowledge of some types of assessment. Therefore, their answers including these misconceptions and lack of knowledge constituted the category of weak literacy. In this context, for example, the answers given by the four prospective teachers about summative assessment are included in this category. Of these, P3-2 and P3-5 stated that the summative assessment is process-oriented, while P3-1 and P3-6 commented that the teacher makes interpretations by considering the frequent results or weaknesses among the group of learners. To exemplify, P3-1 said, it means “making a judgment based on the majority's results”, while P3-6 explained it with such words: “For example, the exam is over. The teacher examined everyone. He makes an evaluation accordingly. For example, it will be about whether the grammar is learned or not”.

Three of the PSETs also lacked of sufficient knowledge about formative assessment. Regarding this, P3-1 and P3-6 did not have any opinions, while P3-2 associated it with the limited response items such as multiple-choice questions and true-false statements. P3-2, who thought that this type of assessment had a more formal structure, based on its name, made the following statement:

I think formative assessment is a bit more stringent. Let me put it this way: opinion cannot be used. Maybe. I have no idea. For example, let me put it this way: we have just given an example. As you know, we have already said that different answers can be given to those open-ended questions that require comprehension that I mentioned here. Not open-ended questions. But for multiple-choice questions or true-false questions, that is, questions where the answer was definitely obvious. They [teachers] would usually add a writing section and give points from there. I think it's wrong. You know. Exams where the answer is clear, if it is 96, it should remain 96. These kinds of exams, these kinds of assessments, I think, can be qualified as formative (Semi-structured Interview III).

As for the performance-based assessment, P3-1 made an incorrect prediction based on its name. She defined it as scoring only on the basis of the student's performance in the assessment without the teacher considering the student's inattendance or opinions about the student in the scoring.

As in the first interview, the results from the third interview showed that none of the pre-service teachers knew what the DA was. Four of these PSETs gave different answers. For example, they defined the DA as an observation made by the teacher during the lesson or during the semester, making changes on teaching according to the feedback from the

students, and assessment done during the in-class activities. P3-5, for example, stated that the DA, like the formative assessment, involves changes made by the teacher according to the feedback from the students:

I can't remember much right now. When I go by the name, the fact that it is dynamic reminds me that it is an assessment that is spread over the process, that is in the process. The reason for being dynamic seems to me that it is done in order to change something and make it dynamic according to the feedback from the student, maybe like formative (Semi-structured Interview III).

The CRA and NRA are two types of assessment about which the teacher candidates in Case 2 need to gain sufficient knowledge. The findings regarding the first of these indicate that the prospective teachers define this type of assessment as the scores assigned to some specific areas like grammar. Moreover, one of the participants (P3-2) explained it as the teacher assigning some extra scores to the assessment results by considering the student's in-class participation.

As for the NRA, only two of the five respondents could comment on it. Of these, P3-2 stated that the NRA has a more restricted scoring system than the CRA and that scores are dichotomously assigned to the students' performances in the NRA. P3-4, on the other hand, compared these two types of assessment and emphasized that the NRA does not involve criteria used as a benchmark; instead, it aims to determine the overall level of the student as in these remarks: "We can think of it as testing in which we evaluate the student individually, not according to a criterion, but we evaluate his/her general situation."

Strong Literacy of the Types of Assessment (Third Semi-structured Interview)

In this section, there are the PSET's correct definitions and explanations about the types of assessment. For instance, three of them accurately defined the summative assessment by emphasizing its product-oriented structure. Similarly, the same PSETs (P3-3, P3-4, P3-5) explained the summative assessment correctly. However, they made references to different features of this type of assessment. To illustrate, P3-3 stated that the formative assessment is process-oriented and that project assignments and portfolios are deployed within this type of assessment. On the other hand, P3-4 stated that this assessment includes feedback that will contribute to the student's learning. Likewise, P3-5 also reported that feedback is vital

in the formative assessment. He also emphasized that this feedback will contribute to the students' learning and enable the teacher to adapt the teaching methods and techniques.

Diagnostic assessment is another type in which most teacher trainees know accurately about what it is. They pointed out that the diagnostic assessment involves the deployment of the methods appropriate to identify the learners' strengths and weaknesses. Regarding this, P3-1 said that, for example, it could be detected that the student's listening skills are developed while his or her writing skills are not developed enough.

As with the diagnostic assessment, the third-grade student teachers were familiar with the performance-based assessment and gave accurate descriptions by referring to the techniques utilized for performance-based assessments. Written and oral assignments were the assessment techniques most mentioned by the prospective teachers. In addition, the prospective teachers mentioned that role-playing, observation, and portfolio are other assessment techniques. As understood from these techniques mentioned, the participants know that the performance-based assessment focuses on how the students use their language skills.

All the PSETs are familiar with the traditional and alternative assessments. Having experienced traditional assessment techniques for years, all teacher candidates mentioned that this assessment is paper-based and that limited response items like multiple-choice questions, true-false statements, and fill-in-the-blanks are frequently used in traditional assessments. Furthermore, they could explain the alternative assessment by emphasizing that it allows more communicative and authentic assessments. Moreover, they referred to compatible methods like project assignments, portfolios, tasks, role-play, and debate.

The definition about the CRA made by P3-1, P3-4, and P3-5 is correct. It involves a reference to the comparison of the students' scores with a benchmark or a specific cut-point score. Regarding this, P3-5 remembered that this was covered in the MEE course and said, "Criterion-based assessment seems to me a bit like absolute evaluation. There are these criteria. You know, there's a criterion of getting 60. If you get 60, you pass. If you get less than 60, you fail." He also knew what the NRA was and defined it as follows: "How good you are within the students. How good or how bad are you in that norm? An assessment to test this."

Awareness of the Factors Affecting the Quality of Language Assessment

The findings regarding the teacher candidates' awareness of the factors that might have an impact on the language assessment are presented in this section. These include their prior assessment experiences, the teacher's qualities to conduct high-quality assessments, and the methods used to teach and learn the target language. According to these findings, the third-year teacher trainees have a high awareness of how these factors might have an impact on the language assessment. Therefore, their answers constitute the category of high literacy.

Table 28

Awareness of the Factors Affecting the Quality of Language Assessment

STRONG LITERACY
<i>The Effects of Assessment Experiences on Language Assessment Approach</i> • Positive effects of positive experiences on assessment knowledge and attitude (P3-3, P3-4, P3-6) • Negative experiences leading to increased awareness (P3-1, P3-2)
<i>The Qualifications Necessary for Teachers to Conduct Quality Assessments</i> • Being competent in language skills (P3-1, P3-3, P3-5) • Paying attention to fairness (P3-2, P3-4, P3-6) • Being a good observer (P3-4, P3-5, P3-6) • Being competent in language assessment (P3-1, P3-5) • Having enough pedagogical knowledge (P3-1, P3-6) • Transparency (P3-2) • Decisiveness (Following the same criteria throughout the whole assessment process) (P3-2) • Knowing grammar and pronunciation (P3-3) • Following the developments in the field of ELT (P3-3) • Being competent in using technology (P3-3) • Being consistent in terms of the assessment content and what is taught in the lesson (P3-4) • Being able to make plans about the assessment process (P3-4) • Providing timely and clear feedback to students (P3-4) • Being competent in test preparation (P3-5) • Knowing the students well enough (P3-5) • Using various assessment methods (P3-6) • Giving importance to ethical and moral values (P3-6)
<i>Theoretical Knowledge about Language Assessment</i> • Knowledge about the principles of assessment (P3-3, P3-5) • The preparation of the questions according to their difficulty level (P3-1) • Information about how to solve the measurement errors (P3-1) • Knowledge about simple statistics (e.g., median) (P3-3) • Knowledge about how to prepare a checklist (P3-4) • Knowledge about how to prepare a test (P3-5) • Knowledge about question types (P3-5)
<i>The Effects of Language Teaching Approach on Language Assessment Practices</i> • The direct reflection of teaching approach on assessment (P3-1, P3-2, P3-3, P3-4, P3-5)

The teacher candidates are aware of how their assessment experiences affected their approach to language assessment. Two important findings emerged in this regard. One of them is that the positive experiences of the teacher candidates had a positive impact on their knowledge and attitudes toward the assessment. The second and more interesting one is that the negative experiences of these PSETs led to an increase in their awareness of assessment. That is to say, as these assessments did not allow them to learn language in a communicative way, they realized what not to do in terms of assessment. For example, regarding this, P3-1 said:

I think in the learning process, in the learning journey, I learned the most from negative examples. The assessment processes of English language skills that I encountered, that is, completely negative until the university. [...] You know, this is how they have measured it so far, but I can't say that from now on you will do it as completely as you can. I can't bring puppets to every lesson. Maybe I can't design toys for every lesson, I can't find different assessment activities for every lesson. But from now on, I received the message that I should emphasize language skills as much as I can and develop such an assessment and evaluation system by taking into account the development of students on these skills (Focus Group Interview I).

As for the qualities that an EFL teacher should have to conduct quality assessments, the pre-service teachers reported that the teacher should use the language skills competently. In addition, having an awareness of the fairness of assessment is another quality that a teacher should have. They also emphasized the importance of being a reasonable observer, proficiency in language skills and assessment, and having adequate pedagogical knowledge.

Four of the senior teacher trainees mentioned that English teachers should have theoretical knowledge of language assessment. In this regard, they emphasized that an English teacher cannot prepare and administer good assessments if s/he does not have sufficient theoretical knowledge. They also said that regarding this theoretical knowledge, one should have knowledge about test and checklist preparation methods, question types, and simple statistics such as mode and median.

Finally, the teacher candidates in Case 2 noted a direct relationship between teaching methods and assessment practices. For example, P3-5 explained how there is an interrelationship between these two elements in the following statement:

An English teacher's approach to teaching and learning English significantly affects assessment practices. Because the teacher needs to assess the objectives related to knowing English, the objectives s/he wants to create in the students, if, for example, for an English teacher, if sitting in a cafe and giving an order is one of his or her objectives, in assessment practices, s/he will test this with a role-play activity in the exam. Two people can measure this. In other words, if their objectives are very good and their objectives are targeting the four language skills, this is the

same. You know, it is reflected in assessment practices in a reflective way (Focus Group Interview I).

Literacy in the Skills Component of LAL

Assessing the Language Skills

The findings related to how the teacher candidates in Case 2 will assess their prospective students' language skills are presented in this section. While they are categorized according to each language skill, they are also detailed in terms of two groups of learners: young learners and adolescents.

Assessing the Listening Skills

Assessing the Listening Skills of Young Language Learners

The findings of this research reveal that the third grade teacher candidates will employ both non-communicative and communicative assessment methods, techniques, or tools to assess the listening skills of their prospective students, regardless of whether they are young learners or adolescents. However, as in Case 1, the answers involving the non-communicative and conventional practices are given under the category of *weak literacy* while the communicative and contemporary practices exist under the category of *strong literacy*.

Table 29

Assessing the Listening Skills of Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P3-1, P3-2, P3-3, P3-5) • Reordering the sentences (P3-4, P3-6) • Filling in a worksheet as an assignment (P3-1) • Matching (Synonyms & Antonyms) (P3-2) • True-False statements (P3-4)	• Listen and spot (P3-1, P3-3, P3-4, P3-5, P3-6) • Listen and act (P3-2, P3-3, P3-5, P3-6) • Listen, draw, and color it (P3-2, P3-3) • Matching (P3-3, P3-4) • Filling in a table with specific information in the audio (P3-3) • Writing the word under the picture after listening to the audio (P3-3) • Listen and chant (P3-4) • Question and answer (P3-4) • Games (P3-4) • Listening to the audio and writing down the basic points (P3-5) • Preparing a poster according to the answers the students got from their peers (P3-5) • Writing a short text according to the answers the students got from their peers (P3-5) • Administering formative assessment methods (P3-5)

-
- Students passing what they heard from peers to other students (P3-5)
 - Listen and show (P3-6)
-

Weak Literacy in Assessing the Listening Skills of Young Language Learners

The teacher trainees in Case 2 reported that they would employ some conventional and non-communicative tools for assessing their future YL students. Fill-in-the-blanks was the most frequently cited item by the participants. They will ask their students to fill in the blanks with appropriate words or grammar structures after they listen to songs or watch cartoons. These items can be given to the students in written examinations, but they can also be given to them in homework assignments in the form of worksheets. Additionally, ordering the parts of the sentences or the separate sentences by listening to audio was another tool that two participants would draw on. P3-4 stated that she would ask the learners to listen to the audio recordings and put the elements of sentences in order to produce grammatically correct sentences. On the other hand, P3-6 reported that in the written exams, the students would be asked to listen to the audio recordings and put the mixed sentences in order to create a meaningful dialogue. Furthermore, true-false statements would also be used by P3-4.

Strong Literacy in Assessing the Listening Skills of Young Language Learners

The participants in this case plan to use more communicative assessment methods when assessing the listening skills of their future primary and lower secondary school students. Listen and act, and listen and spot, are two practices that most participants intend to use. Emphasizing that the physical activities are more suitable in consideration of the characteristics of the group of learners, the third graders also frequently referred to the total physical response (TPR) as a teaching method. To illustrate, P3-5 made the following explanation for assessing the primary school students:

In primary education, since TPR is recommended in the curriculum, it can be a listen and do assessment. Listen to me, or go get the red apple. Listen to me; go open the window and open the door. We do something like determine whether the student is able to fulfill the instruction they hear at that moment. Because it starts in 2nd grade. It doesn't seem logical and realistic to me to make an evaluation with long listening pieces in the evaluation part of the student, except for such input. So, for more TPR activities, maybe songs can also be used here. In some songs, clap when you hear blue, clap when you hear red, and jump. The student can do something kinesthetically (Semi-structured Interview III).

Listen and spot, requiring the YLs to detect the objects after listening to short instructions or descriptions, will be used by the four student teachers. Regarding this practice, the participants stressed that the materials or objects they were familiar with would be used. For example, according to P3-2, the objects in the classroom would be useful for this assessment practice.

Listen, draw, and paint is another assessment practice that two pre-service teachers in Case 2 will benefit from. As the reason for adopting it, they had the opinion that drawing and coloring would be more appropriate for assessing, especially the primary school students since they were not proficient enough to write long sentences and drawing and coloring would not cause stress on the students during the assessment. Noting that the fill-in-the-blanks items would be challenging for the primary school students, P3-2 expressed that he would ask his students to watch a section of a cartoon movie and draw an object and color it according to the description given in it without seeing how that object was in the movie.

Gap-filling, actually. I mean, apart from gap filling, I can't think of much right now. I think I'm having a little more difficulty here because I don't listen, but how can it be evaluated? Gap filling may be difficult for children in primary school. But maybe it can be like this: I can ask them to draw something based on what they hear. For example, in the cartoon, the apple is described in very simple language. I can ask the children to draw an apple. I mean, let me see if they understand what they are listening to. For example, the leaf is this color, and the stem is this color. I change this, and I definitely make them draw pictures in primary school. Because children cannot learn to write in the second grade, especially in this period, it may be difficult for them to fill in too many blanks (Semi-structured Interview III).

Assessing the Listening Skills of Adolescent Language Learners

As with the YLs, the teacher trainees in Case 2 will employ both non-communicative and communicative techniques and tools to assess their adolescent students. In other words, they will ask closed-ended questions including fill-in-the-blanks, true-false, matching and multiple-choice questions, as in the table below. On the other hand, they will also use assessment methods that integrate their adolescent students' speaking or writing skills with listening.

Table 30

Assessing the Listening Skills of Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P3-1, P3-2, P3-3, P3-5) • True-False statements (P3-1, P3-4, P3-5) • Matching the words with their synonyms or antonyms after listening to a text (P3-1, P3-2) • Multiple-choice questions (P3-2, P3-5) • Short-answer questions (P3-4)	<i>Listening and speaking integration</i> • Discussing and comparing their city with one in audio (In pairs) (P3-3) • Listening to a friend, filling in a table, and speaking (P3-3) • Watching a video, taking notes, and question-and-answer (P3-4) <i>Listening and writing integration</i> ◦ Note-taking (P3-1, P3-3, P3-4, P3-5) ◦ Writing a summary of the audio (P3-1) ◦ Open-ended questions (P3-5) • Asking the main idea of the audio (P3-4, P3-5, P3-6) • Filling in a table or mind map with specific information in the audio (P3-3, P3-6) • Listening to the audio and finding some details (P3-4) • Ordering the given pictures according to the audio (P3-6)

Weak Literacy in Assessing the Listening Skills of Adolescent Language Learners

In order to assess their adolescent students, the participants will use some traditional methods. Fill-in-the-blanks, true-false statements, multiple-choice questions, and matching items are the types of assessment tools that most of the student teachers plan to use. While the participants usually stated that they would administer these items in a pen-and-paper format, P3-2 pointed out that he would also use Kahoot for listening assessments and answer the fill-in-the-blanks and multiple-choice questions on this platform while listening to popular songs or afterwards. Moreover, two participants (P3-1 and P3-2) will assess their students' listening skills through matching items. In these, the learners will match the given words with their synonyms and antonyms after listening to songs or watching a documentary.

Strong Literacy in Assessing the Listening Skills of Adolescent Language Learners

In contrast to the non-communicative assessment tools mentioned above, the participants will also use methods enabling their students to integrate listening with speaking or writing skills, or both, and use their listening skills in a way that is appropriate for their use in real life. For example, P3-3 wrote in the scenarios that she would observe her students while they

were talking in pairs and comparing their city with the one described in the audio. In this assessment, the learners will be required to both comprehend the information from the listening text and interact with their peers by speaking English. The assessment that will be carried out by P3-4 will include the integration of listening with both writing and speaking skills. In this vein, the learners will take notes while watching a video and then answer the questions asked by the teacher. Like P3-4, three other PSETs (P3-1, P3-3, and P3-5) will also ask their prospective adolescent students to take notes. Moreover, P3-1 stated that her students would write a summary of the listening text based on the notes they had taken. As for P3-5, he will ask his students open-ended questions that they will answer based on the main idea and details in the listening text.

The teacher candidates will also adopt other techniques that require the learners to comprehend the gist and details of the audio. For example, P3-4, P3-5, and P3-6 will ask their students what the main idea of the listening text is. Moreover, P3-3 and P3-6 will also ask to fill in a table or a mind map with the specific information given in the audio. Regarding this, P3-6 wrote in the scenarios: “I prepare questions appropriate to the listening text. For example, ‘gist and specific questions’. Firstly, I ask what the topic is in the first listening, and then I ask various mind-map-style questions that ask for detailed information.” As for P3-5, he will ask his students written open-ended questions that they will answer based on the main idea and details in the listening text.

Assessing the Speaking Skills

The findings of the research reveal that the LAL level of the PSETs in the third year is high in terms of assessing the speaking skills of YLs and adolescents. In other words, all of them aim to use communicative and authentic methods in the speaking assessments that they will conduct in the future. Therefore, the following sections present only the methods or practices that are categorized as indicators of strong literacy in the assessment of speaking skills.

Assessing the Speaking Skills of Young Language Learners

The PSETs in Case 2 referred to various methods compatible with the communicative language assessment approach (Brown & Abeywickrama, 2018, pp. 15-16), as given in the table below. These will enable the teacher to directly assess their young students’ speaking

skills while they are interacting with their peers or teacher or making individual speeches. Moreover, it should also be reported that these methods will be used in the context of both formal and informal or alternative assessments.

Table 31

Assessing the Speaking Skills of Young Language Learners

STRONG LITERACY
• Role-play (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6) • Presentation (P3-1, P3-3, P3-4, P3-5) • Question-and-answer (P3-1, P3-2, P3-4, P3-5) • Student-student conversation (P3-1, P3-2, P3-3, P3-6) • Recording short videos (P3-1, P3-4)

Role-play is an assessment method that all the participants intend to use for their future speaking assessments. In their explanations regarding this method, they emphasized that they would expect the learners to communicate with each other with simple and short sentences. Besides this, some of them noted that role-play cards and scenarios would be given to them in order to create a communicative context. For the primary school students, they would take this as an informal assessment, and the teacher would use it to collect data about their students' speaking skills and give feedback on their errors. On the other hand, they will ask their students to role-play as a part of formal assessments and will assign scores to their speaking performance.

Making a short presentation is another frequently mentioned method. This presentation will be about simple and familiar topics, like the introduction of family members. Moreover, they will use the posters or drawings that they will prepare before the presentation. Regarding this, P3-2 wrote:

Beforehand, I would distribute an activity for drawing and painting to the students in which they could show the physical characteristics of their families. After completing this activity, I would ask each student to hang their work on the board and make a short presentation about their family, including their physical characteristics and abilities, through visuals. While making my assessment, I would pay attention to whether the student used the target words correctly in the sentence in his speech. For example, it is important for me to use the expression 'blonde' for her mother, whom she paints as blonde, and I will give my score there, but making a grammatical error like my mother *have* blonde hair while making a sentence would not affect my score, but I would correct this mistake (Scenarios for Foreign Language Assessment).

The PSETs will also have short question-and-answer conversations with their students to assess their speaking skills. These conversations will consist of the teacher asking some

simple questions and the learner answering them. For example, their questions will be about their names, ages, favorite colors, and toys, as illustrated in the excerpt below.

In primary school, again, we can evaluate their answers to questions without expecting anything too big, with small things like "What is this?" "This is an apple" or a very simple "What is your name?" "That", "This," or "What is this?" in a conversation. But here, I think that in-class observations may be more logical, not with the logic of an exam, because at the primary school level, I guess they don't do much to associate this grading with a number. Since there is a very good, good, kind of evaluation style, we can base speaking skills more on observations in the classroom and not so much on grading (P3-5, Semi-structured Interview III).

In addition to the conversations between teacher and student, they will also create and administer assessments consisting of conversations between or among the students. In these conversations, they will introduce themselves or their family members to their classmates or discuss a topic. Here, it should be noted that although role-play also involves communication between students, it is performed in the context of specific scenarios and the roles given on the role cards. On the other hand, the student-student conversations are a bit more open-ended and enable the learners to express their opinions.

Asking the students to record short videos is another method that two of the third-year PSETs will employ. This method, which will be applied by P3-1 and P3-4, includes the students video-recording their own short speech about a topic assigned by the teacher or one they choose. Regarding this method, the participants expressed that it would prevent them from limiting their speaking assessments to the classroom.

Assessing the Speaking Skills of Adolescent Language Learners

The PSETs in Case 2 will also utilize communicative assessment methods to assess the speaking skills of the adolescent or upper secondary school learners of English. As given in the following table, they consist of methods requiring the students to interact with each other, with the teacher, or to speak individually. In this regard, these methods are categorized as indicators of the participants' high level of literacy in assessing the speaking skills of adolescents.

Table 32

Assessing the Speaking Skills of Adolescent Language Learners

STRONG LITERACY
• Role-play (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6) • Pair or group discussion (P3-1, P3-3, P3-4, P3-5) • Making a presentation (P3-1, P3-3, P3-4, P3-6) • Question-and-answer (P3-2, P3-4) • Drama (Listening/Reading + Speaking) (P3-3) • Listening to a friend, filling in a table, and speaking (Listening, Writing, Speaking) (P3-3) • Discussing a text and filling in a table (Reading, Speaking, and Writing) (P3-3) • Groupwork activity for problem solving and reasoning (P3-4) • Individual interviews (P3-4) • Talking in pairs (P3-4) • Preparing videos (P3-4) • Preparing digital stories voiced by the students (P3-4)

Role-play is the method that the majority of the participants prefer to employ in the speaking assessments they will administer in the future. They explained in the journals, interviews, or scenarios that they would use this method of assessment. To give an example, P3-1 wrote in her Opinion Journal IV that she would carry out speaking assessments “both through improvisation and through pre-prepared scenarios”. Also, she noted that the learners would be expected to use not only formal but also informal expressions according to the context of the role-play. In the scenarios, P3-2 wrote that he would ask his students to act out the roles he would give them. In the interview, P3-3 stated that she would integrate listening or reading with the speaking skills and would ask them to role-play by drawing on the content given in the listening and reading texts. Moreover, she said in the interview that she would set aside a separate day for the assessment of speaking skills with various methods, and the role-play would be one of them. In the scenarios, P3-5 reported: “Since the students are expected to retell and enact the story and talk about past habits or traditions, I use role-play and drama activities to assess speaking skills”. During these activities, he will videotape the students’ performances “as long as the conditions permit,” with the aim of watching them later and ensuring fairness. P3-6 made a reference to the activities she used for the microteachings in the Teaching English Language Skills courses and put an emphasis on the use of the prompts for providing a simulation of real-life situations.

The adolescent students will also make presentations for the speaking assessments. The assessments that will be conducted by P3-1 will include presentations about a book that the students have read or past habits and traditions. Likewise, P3-3 also gave the answer in the

scenarios that “I would ask students to prepare a presentation individually and talk about their past habits and traditions.” In the interview and journal, P3-4 wrote that the presentations her students would make would be about a topic they chose or the poster they would prepare comparing their traditions and the current situation of them. P3-6 will also ask her students to tell a story or talk about past habits and traditions in their presentations. Adolescent students will also discuss in pairs or groups as a part of the speaking assessments. In this regard, P3-1 wrote in her journal that “[I] can conduct assessments based on interaction” and “I can assess students in pairs or groups for speaking skills”. In line with this thought, she drew on her experiences in the courses at the university and said that her students would discuss in pairs or groups and try to find solutions to a problem or generate new ideas. P3-3 wrote in her journal and stated in the interview that she would conduct group or pair discussions. In the interview, she emphasized that the assessment should not create stress on the students, and therefore she would assess her students’ speaking skills with the in-class discussion activities. In these discussions, she will expect her students to generate new ideas, and she will take notes by monitoring their performances. Similarly, P3-4 reported that her students would discuss to find solutions to the given problems and detect their reasons. Moreover, she wrote that “I want my students to make self-assessments and peer assessments” for each language skill. Therefore, she will enable them to reflect on their own and their friends’ performances. In fact, this is compatible with their experience in the courses Teaching Language Skills I and II because the professor of the course expects the teacher candidates to have a self-assessment section in their lesson plans and micro-teachings.

Assessing the Reading Skills

Assessing the Reading Skills of Young Language Learners

The third-year teacher candidates will use not only non-communicative or conventional assessment tools but also communicative and authentic techniques to assess the reading skills of their YLs. However, it should be noted that the findings in this section only relate to the assessment of lower secondary school students’ reading skills since these skills are not given a major focus in the primary school curriculum.

Table 33

Assessing the Reading Skills of Young Language Learners

WEAK LITERACY	STRONG LITERACY
• True-false statements (P3-4, P3-5, P3-6) • Fill-in-the-blanks (P3-3, P3-4) • Multiple-choice questions (P3-4)	• Open-ended questions (P3-2, P3-3, P3-4, P3-5) • Writing a summary or a review (P3-1) • Reading a ticket, discussion, and role-play (P3-2) • Matching the short texts with the pictures (P3-3) • Ordering the mixed parts of a text (Group Work) (P3-3) • Reading and making a presentation (P3-5) • Filling out a mind map (P3-6)

Weak Literacy in Assessing the Reading Skills of Young Language Learners

Some PSETs in Case 2 will use mechanical and non-communicative assessment tools such as true-false statements, fill-in-the-blanks, and multiple-choice questions. Of these, true-false statements will be employed by P3-4, P3-5, and P3-6. While P3-4 said that “I can use short but understandable texts” like short stories and that she would include true-false statements in her reading assessments, P3-5 made the following explanation in the third interview:

In secondary school, considering that students have mastered reading and writing skills in their own language, we can introduce reading texts here a little bit. But it should not be too long, and it can be done by paying attention to the level of the students. [...] At this stage, true-false statements can be put so that the student can understand them a little more easily (Semi-structured Interview III).

P3-6, who is another participant planning to use true-false statements, expressed that she examined her brother’s English coursebook and realized that short dialogs and paragraphs are used in that book. Therefore, she pointed out that she would give such short texts to her students and include true-false statements to assess reading skills in the written exams.

Fill-in-the-blanks will be another item that two PSETs will employ in their reading assessments. For example, P3-3, who was one of these teacher candidates, said, “Reading is one of the skills I have the most difficulty with,” and therefore, she was not sure how to assess the reading skills of YLs. In this context, however, she stated that she would draw on short dialogs and fill-in-the-blank items in her reading assessments. Stating that she will ask

various types of items for assessing her students' reading skills, P3-4 will also give short dialogs to her students and ask fill-in-the-blanks and multiple-choice items.

Strong Literacy in Assessing the Reading Skills of Young Language Learners

The teacher candidates in Case 2 will also utilize more communicative or authentic assessment methods or tools in contrast to the ones mentioned in the previous section. Open-ended questions will be the most preferred tool for assessing the comprehension of the reading texts. These questions will be asked within the formal and informal assessments, and the learners are expected to write or verbally express their answers to these questions by making inferences from the text. Moreover, they will be about such texts as short paragraphs, dialogs, and stories. P3-1 and P3-2 will integrate reading and writing skills. Of these, P3-1 will ask her students to write a summary or a review of a short paragraph, dialog, or book, while P3-2 will conduct an assessment in which the learners will fill in a mind map after reading the text. On the other hand, P3-2 and P3-5 will aim to integrate reading and speaking into their assessments. With this aim in mind, P3-2 will give some scenarios and tickets to his students, and the students will role-play as a ticket seller and a customer according to the scenarios and the tickets they have. Likewise, P3-5 wrote, "I plan to do assessments where students can provide an output with what they understand from the text". In this respect, he will give short reading texts to his students and then ask them to make a short presentation about what they understand from the text.

Matching will be a tool that P3-3 will ask her students to answer. In this assessment, the learners are to match the pictures with short sentences. Moreover, for carrying out an informal assessment, she will divide a short text into 3 or 5 pieces and then give her students these parts in a scrambled order and ask them to work in groups to put these parts in order.

Assessing the Reading Skills of Adolescent Language Learners

Both non-communicative and communicative assessment methods or tools will be employed by the third-year teacher trainees for assessing the adolescent or upper secondary school students' reading skills, as shown in the following table. The non-communicative tools include true-false statements, fill-in-the-blanks, and multiple-choice questions, and they will

measure only the learners' ability to comprehend the text with a purpose incompatible with the use of reading skills in real-life. On the other hand, communicative or authentic assessment methods or tools include the integration of reading with speaking or writing skills or are in a format of more free response.

Table 34

Assessing the Reading Skills of Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• True-false statements (P3-1, P3-2, P3-4, P3-6) • Fill-in-the-blanks (P3-1, P3-3, P3-4) • Multiple-choice questions (P3-4)	<i>Reading and Speaking Integration</i> • Reading a ticket, discussion, and role-play (P3-2) • Reading and discussing a text and filling in a table (P3-3) • Reading and making a presentation (P3-5) <i>Reading and Writing Integration</i> • Open-ended questions (P3-2, P3-3, P3-4) • Writing a summary (P3-1, P3-2, P3-4) • Writing a review (P3-1, P3-4) • Completing a half-written text (e.g., a story) (P3-3) • Writing short paragraphs or essays after reading about a topic (P3-5) <i>Other</i> • Matching (P3-1, P3-2) • Finding the main idea or details of the text (P3-4, P3-6) • Filling in a table by using information in the text (P3-3) • Reading a text and filling in the table with suggestions (P3-3) • Ordering the mixed parts of a text (Group Work) (P3-3) • Ordering the events in a tabular form after reading the text (P3-5) • Completing the map with details in the text (P3-6)

Weak Literacy in Assessing the Reading Skills of Adolescent Language Learners

In this section, limited response items like true-false statements, fill-in-the-blanks, and multiple-choice questions will be included in the written exams that the pre-service teachers will administer to their adolescent students. They mentioned these types of items in the journal, interview, and scenarios. For example, P3-1 wrote in the scenarios that she would include it in the written exam on reading skills, and the reading text would be the biography of a famous person. Similarly, P3-2 gave the answer in the scenarios that he would administer a written exam and ask true-false statements as well as open-ended questions and matching. P3-4 reported both in the journal and in the scenarios that she would ask true-false questions in a written exam, which would also include limited response items like multiple-choice questions and fill-in-the-blanks. Additionally, she noted that “when assessing reading

skills, I pay attention to the authenticity of the materials I use”. In contrast to these written references to the true-false questions, P3-6 mentioned their use in the interview and pointed out that she would utilize them in addition to the more communicative assessment methods.

*Strong Literacy in Assessing the Reading Skills of
Adolescent Language Learners*

Assessment of the reading skills of adolescent learners will also be done through more communicative and authentic methods or constructed-response items. In other words, these methods include the integration of reading with speaking or writing skills. Regarding the integration of reading and speaking, the participants pointed out in a written and verbal way that they would give texts of different genres to their students, such as dialogs, short stories, and song lyrics, and ask them to discuss, role-play, or make short presentations about these texts.

As for the integration of reading and writing skills, it involves the students reading about a topic and writing answers to open-ended questions or a piece of text based on what they understand from the texts. For example, P3-2, P3-3, and P3-4 will ask open-ended questions, requiring the learners to make inferences from the text to give answers. Moreover, P3-1, P3-2, and P3-4 will ask their students to write a summary or a review after reading a text. The texts that will be summarized or reviewed will be lyrics, posters, film reviews, and books. P3-3 also plans to integrate reading and writing skills into her assessment, as explained in the excerpt below:

Story completion, which attracts my attention the most and which I think will be efficient, can be a good assessment tool. When I give the beginning of an interesting story and ask the student to bring the continuation, I think it will be much more productive than if the student writes a story or text from scratch. In addition, if I support the beginning of this story with visuals according to the order of occurrence, it will be very useful for students (Opinion Journal IV).

P3-5 pointed out that students are in adolescence and that using reading passages related to their interests will motivate them more. In this regard, he said that he would give longer and more advanced texts to his students compared to the lower secondary school students and that he would ask his students to write short paragraphs or essays if their language level was high enough.

Assessing the Writing Skills

Assessing the Writing Skills of Young Language Learners

The PSETs in Case 2 plan to exploit authentic or real-like tasks or techniques to assess their prospective young students. Through these, they will be able to directly analyze the students' writing skills. Therefore, the findings in Table 35 reveal that the participants have high literacy in assessing the writing skills of YLs in English.

They will ask their students to write various types of texts in their writing assessments. To illustrate, P3-1, P3-3, and P3-4 will conduct a process-oriented assessment, and their students will write short diaries as a part of this assessment. While the pupils of P3-1 and P3-4 will write their diaries in a pen-and-paper format, the ones of P3-3 will use a digital platform to do so.

The third-year teacher candidates would also carry out more product-oriented assessments. Of these, writing a short description about a favorite activity or an object, drawing a picture, and preparing a poster will be the assessment techniques that P3-1 will exploit. As for P3-3, her students will write a short comment under a given picture. P3-4, planning to use various assessment techniques like self- and peer assessment, explained what she would employ as follows:

In lower secondary school, there may be things like filling in tables. There may be things like preparing a post card. Or they can start doing something like writing short letters in secondary school from the 6th grade onward. They can start doing things like filling out small job application forms or CVs and introducing themselves in secondary school (Semi-structured Interview III).

Table 35

Assessing the Writing Skills of Young Language Learners

STRONG LITERACY
• Writing a short text like a diary, paragraph, e-mail, letter, or postcard (P3-1, P3-4, P3-5) • Assessing spelling (P3-5, P3-6) • Drawing a picture and describing it (P3-1) • Preparing a poster (P3-1) • Writing a short comment under the given picture (P3-3) • Filling in a table, job application form, or CV (P3-4) • Self-assessment (P3-4) • Peer-assessment (P3-4) • Writing answers to open-ended questions after reading (P3-5)

For speaking and writing skills, P3-5 wrote in the journal “First I think to give the student a purpose and provide input”. As the reason for that, he gave the following explanation: “When measuring these skills, if the student understands the purpose very well and gets the

necessary input, I think it will provide much better results”. In line with this thought, he reported that “[he] would use situations that are in life and that students may encounter in order to assess [writing skills] in a communicative way”. For this purpose, he will ask his students to write e-mails and petitions.

Assessing the Writing Skills of Adolescent Language Learners

The students studying at an upper secondary school are referred to as adolescents in this research. In this respect, the writing skills of the adolescents will be assessed through various communicative and authentic tasks by the third-year PSETs. As the most frequent of them, they will assess their adolescent students by asking them to write texts of different genres.

Table 36

Assessing the Writing Skills of Adolescent Language Learners

STRONG LITERACY	
• Writing texts of different genres ○ Complaint letter (P3-1, P3-2, P3-3, P3-4, P3-5) ○ E-mail (P3-1, P3-3, P3-4, P3-5) ○ Review (P3-1, P3-2, P3-3) ○ Essay (P3-4, P3-5, P3-6) ○ Paragraphs (P3-4, P3-5) ○ Summary (P3-1, P3-4) ○ Tweet (P3-2) ○ Address description (P3-2) ○ Online diary (P3-3) ○ Social media comment (P3-3) ○ Blog post (P3-4) ○ Post-card (P3-4) ○ Birthday card (P3-4) ○ Petition (P3-5) • Preparing a poster, brochure, or billboard (P3-1, P3-3, P3-4) • Completing a half-given story (P3-3, P3-4, P3-6) • Answering comprehension checking questions (P3-2, P3-3) • Asking the student to do a search and write (P3-1) • Listening to audio and writing (P3-1) • Portfolio (P3-4) • Creative writing activities (P3-4) • Self-assessment (P3-4) • Peer-assessment (P3-4) • Writing to convey news, stories, and past events (P3-5) • Writing a text according to the given scenario (P3-6)	

A complaint letter is the most commonly cited type of text that the five PSETs’ adolescent learners will write. To enable them to write the letter, the teacher candidates will provide the students with scenarios as prompts. In addition to this, writing an email will be another

authentic task that will be exploited for the writing assessments. Regarding this, P3-3 articulated that administering a writing exam “does not make much sense” and therefore she would assess her students’ writing skills through assignments. As one of these assignments, she will ask the learners to write a reply to an e-mail written by an imaginary friend. In addition, other types of texts such as paragraphs, essays, reviews, summaries, blog posts, diaries, or tweets will be written by the adolescent learners of English. Of these, writing tweets appropriate to the given situation was a task that would be offered by P3-2, and he also used it for teaching writing in his micro-teaching he did within the scope of the TLS II course. Similarly, P3-3 and P3-4 will also relate the forms of their assessments to digital technologies. That is to say, they will write online diaries, social media comments, and blog posts.

Preparing a poster, brochure, or billboard will be a part of the assessments to be conducted by three PSETs. As one of these teacher trainees, P3-1 wrote in the journal that “when making an assessment, we can include many skills in the process.” In line with this, her assessment will integrate reading and writing skills and include the students preparing a poster about a book they read. In the interview, P3-3 referred to what she wrote in the journal and pointed out that she would ask her prospective students to create a poster or brochure using applications such as Canva. P3-4 also mentioned her answer in the journal and repeated it in the interview that her students would make an advertisement billboard through some online platforms like Canva. Moreover, she answered in the scenarios that the adolescent learners would “prepare a poster for the letter of complaint”.

Assessing the Language Components: Young Language Learners

The assessment of language components is another area in which teacher candidates need to be literate. Therefore, the findings regarding how they will assess their students’ use of grammar and vocabulary will be presented in this section. As in Case 1, techniques that involve non-communicative, explicit, and mechanical assessment of students' knowledge of language components are classified under the category of weak literacy indicators. On the other hand, techniques that involve the implicit and communicative assessment of language components are considered indicators of strong literacy.

Table 37

Assessing Grammar: Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P3-1, P3-3, P3-4, P3-6) • Worksheets (P3-1) • Asking the students to correct grammatical errors in a given text (P3-3)	<i>Assessing grammar in speaking</i> (P3-1, P3-2, P3-4, P3-5, P3-6) ○ Peer talk (P3-1) ○ Giving instructions to peers (P3-2) ○ Making a short speech about family members (P3-2) ○ Making sentences with like and dislike (P3-2) ○ Observing whether the student can make simple grammatical sentences (P3-4) ○ Listen or read and speak (e.g., tell me when they will meet) (P3-5) ○ Making sentences by using the target structure according to the given emojis (P3-6) <i>Assessing grammar in writing</i> (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6) ○ Writing a diary (P3-1) ○ Writing a sentence by using the word on the wheel (P3-2) ○ Writing answers to questions according to given pictures and clues (P3-3) ○ Preparing a post-card (P3-4) ○ Writing a short letter (P3-4) ○ Filling out a CV/job application form (P3-4) ○ Writing answers to open-ended questions (P3-5) ○ Paraphrasing the given sentence (P3-6) ○ Completing a dialogue (P3-6)

The indicators of weak literacy involve such mechanical and non-communicative items as fill-in-the-blanks and correcting the given grammatical errors in a text. These items will be administered to the participants in the format of either written exams or worksheets. However, it is important to note that the teacher candidates are intended to teach and assess the language components implicitly, although they will sometimes exploit such mechanical tools under some conditions, as explained in the following excerpt:

In general, I honestly prefer not to teach grammar directly. In the same way, I would like to assess it within other skills. But again, as I said, if it gets mixed in with other skills and [implicit assessment] is not efficient, when I feel that they are not learning, I can give them a text and ask them to correct their grammar mistakes there, for example, but not in a very mechanical way, that is, not directly, for example, choose from the options (P3-3, Semi-structured Interview III).

Moreover, P3-6 said that one of her group mates asked his or her imaginary students to write "can" or "can't" according to the given emoji in micro-teaching and said that this could also be used in grammar assessments.

The participants in Case 2 will also conduct implicit assessments on the use of grammar by YLs. These assessments will be based on written or oral products. To illustrate, P3-2 will focus on how the students are able to use grammatical structures when giving verbal instructions to their peers. Moreover, P3-2 will also use students' talks about family members and their likes or dislikes to assess grammar. P3-5 also emphasized that language components should be assessed implicitly and explained in his diary how to do this assessment:

While evaluating grammar, I plan to look at whether the student can use it rather than knowing the grammar rule. Rather than isolated grammar tests, assessments integrated into language skills would be more beneficial. (e.g., listen or read and tell me when they are going to meet?) With the activity in the example I gave, children will be able to find the answer by focusing on both sentence structures and meaning by focusing on the times in a text they have listened to or read without realizing it (Opinion Journal IV).

In addition to assessing grammar in speaking, P3-6 will also pay attention to how students use grammar in written products. For this purpose, she will ask students to paraphrase a given sentence and complete a dialogue appropriately.

As for the assessment of how YLs of English are able to comprehend and use the lexical items of the target language, the findings reveal that the teacher candidates will draw on both explicit and implicit assessment techniques. In other words, they will employ techniques that will not require the learners to comprehend or use the target vocabulary and are not very compatible with real-life situations.

Table 38

Assessing Vocabulary: Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P3-1, P3-2, P3-3, P3-6) • Matching words with synonyms or antonyms (P3-1, P3-2) • Short vocabulary quizzes (P3-4) • Vocabulary knowledge: More important than vocabulary use (P3-5)	• Assessing vocabulary in speech (P3-1, P3-3, P3-4, P3-5, P3-6) • Assessing vocabulary in writing (P3-2, P3-4, P3-5, P3-6) • Writing a word related to the drawing or photo (P3-1, P3-2, P3-3, P3-6) • Listen and show (P3-2, P3-5, P3-6) • Matching pictures with speech bubbles (P3-3, P3-5) • Listen and paint (P3-2) • Game on Wordwall (Placing the word in a chart) (P3-3) • Taking the picture out of the box and telling the equivalent word (P3-3) • Vocabulary games (P3-4) • Creating a mind map (P3-4)

Fill-in-the-blanks and matching, which are categorized as indicators of weak literacy, will be utilized for assessing the vocabulary knowledge of primary and lower secondary school students. For example, P3-3 reported in the scenarios that she would carry out an assessment in which the students would write appropriate words to fill in the gaps according to the pictures of facial expressions. P3-6 will give written sentences to her students and ask them to write the appropriate word according to the context. Furthermore, the vocabulary assessments will include sections in which the YLs will listen to a text of their interest and match the words with their synonyms or antonyms.

In contrast to these tools of explicit vocabulary assessment, the teacher trainees reported that they would also do implicit vocabulary assessments on the written and verbal products of their students. That is to say, they will pay attention to how the learners are able to comprehend or use the words of the target language through listening, speaking, reading, and writing skills. In this context, three teacher candidates reported that in order to assess the extent to which the students understand the target language vocabulary, they would ask the students to listen to the instructions and perform actions that show their understanding of the target vocabulary. To illustrate, P3-2 stated the following words:

For example, it goes like this: I look at the child's association. When I say yellow, if the child can paint yellow, show the yellow pencil or the yellow card. Or, let's say I'm evaluating words in the exam by listening. If he can paint yellow when I say yellow, the child understands the meaning of that word for me. Mine will be more of a demonstration (Semi-structured Interview III).

Additionally, some PSETs will use the students' speech as the main focus of their vocabulary assessments. P3-3, for example, wrote in the scenarios that she would pay attention to how her students used vocabulary in their presentation to introduce their family members. Unlike this, P3-6 will establish a conversation by asking the student a question and getting an answer, and she will assess how the student uses the vocabulary in this conversation.

Assessing the Language Components: Adolescent Language Learners

Some of the teacher candidates will exploit classical and mechanical tools to conduct explicit assessments of adolescent learners' grammar knowledge. On the other hand, they will also assess the use of grammar through their oral and written products. In this respect, the first group of assessment methods is categorized as weak literacy, while the second group, which allows for more implicit assessment, is categorized as strong literacy.

Table 39

Assessing Grammar: Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Grammar quiz (P3-3, P3-4) • Ordering the scrambled words to make grammatical sentences (P3-3) • Asking the students to correct grammatical errors in a given text (P3-3) • Assessing grammar with online applications or games (P3-4) • Assessing grammar explicitly (Language majors) (P3-5)	• <i>Assessing grammar in speaking</i> (P3-1, P3-2, P3-4, P3-5, P3-6) ◦ Making a speech (e.g., favorite tourist destination) (P3-1, P3-2, P3-4, P3-6) ◦ Role-play (P3-1, P3-2, P3-6) ◦ Speaking activity done after listening (P3-5) • <i>Assessing grammar in writing</i> (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6) ◦ Writing paragraphs (P3-1, P3-5) ◦ Writing answers to open-ended questions related to the reading text (P3-3, P3-6) ◦ Writing simple sentences (P3-2) ◦ Writing a letter (P3-2) ◦ Completing the dialogs in the exam (P3-2) ◦ Writing a blog post using information s/he got through an interview (P3-4) ◦ Writing a text after listening (P3-5)

P3-3 will administer both pen-and-paper exams and online quizzes or games. In the written exams, she will ask her students to correct given grammatical errors and to make grammatical sentences by putting the mixed components of the sentences in the correct order. On the other hand, she will utilize fill-in-the-blank items in the online quizzes. These will consist of ready-made items requiring the test takers to fill in the blanks with appropriate grammatical structures according to the given pictures within the specified time. Likewise, P3-4 will sometimes carry out explicit grammar assessments online, and they will include multiple-choice questions administered through platforms such as Quizlet and Kahoot.

As compatible with the negative washback effect of the university entrance examination, P3-5 stated in the interview that he would conduct explicit grammar assessments for the language majors in the upper secondary school by emphasizing that “the exam is the boss of education,” while he would prefer implicit assessments for other groups of learners. Moreover, while giving this answer, he reminded himself that he was prepared by his teacher for the exam in the same way, and then he got the consciousness that the language components should be assessed implicitly as follows:

So it is the same at the upper secondary school level. Instead of measuring isolated grammar, [it will be] within the students' skills. Of course, we may also need to protect here. If there is a language class in the upper secondary school, unfortunately, we now have to measure it in a deductive way. Because what we always hear is, "The exam is the boss of education". If I am asked a question like "What will come here, what will come here?" in a mechanical way, I have

to prepare the student for this. We were prepared in such a way, and then we reached this consciousness. At such a point, we may have to do the same for the students who will take the language class or the language exam in order for them to be able to get into the English language teaching department (Semi-structured Interview III).

On the other hand, the teacher candidates will also carry out implicit assessments of grammar by paying attention to the adolescent learners' ability to use the target language grammar. In this vein, the speeches that the learners will make for production or interaction will provide the necessary context for the grammar assessments. To be more precise, five of the PSETs will deploy various speaking tasks like role-plays, giving an individual speech about a topic, or speaking after listening to audio.

Writing will be the other language skill through which the teacher candidates will assess their prospective students' performance in making grammatical sentences. Emphasized by all participants, this skill will be elicited through tasks such as writing paragraphs, essays, letters, dialogs, blog posts, or answers to the open-ended questions. Here, it should be noted that the adolescents will both comprehend and use grammar through some speaking and writing tasks, which will enable the PSETs to implicitly assess their grammatical knowledge and use.

As for the assessment of the vocabulary knowledge and use of the adolescents, the members of Case 2 will utilize both traditional or mechanical tools for explicit or discrete-point assessment and authentic or communicative tasks for implicit assessment. Fill-in-the-blanks, matching the words with synonyms or antonyms, writing the Turkish equivalents of English words, and multiple-choice questions will be included in the vocabulary tests that will be administered to measure the vocabulary knowledge of the students in the format of written exams or online quizzes.

Table 40

Assessing Vocabulary: Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P3-2, P3-3) • Matching the word with an antonym, synonym, or homophone (P3-2) • Taking three pictures out of the box and writing their Turkish equivalents (P3-3) • Digital puzzles (P3-4) • Digital quizzes (P3-4) • Multiple-choice questions (P3-6)	<i>Assessing Vocabulary in Speaking</i> • Role-play (P3-1, P3-3, P3-4, P3-6) • Short speech about a given topic (P3-1) • Making a presentation (P3-4) • Speaking activities done after listening (P3-5) • Reading and speaking (P3-6) <i>Assessing Vocabulary in Writing</i> • Writing a paragraph (P3-1, P3-5) • Describing a given artwork (P3-1) • Preparing a poster (P3-1) • Writing activity done after listening (P3-5) • Guessing the meanings of the words in bold and writing a sentence with them (P3-5)

-
- Answering open-ended questions (P3-6)
- Games*
- Game on the Wordwall (Matching the word with the related picture) (P3-3)
 - Game on the Wordwall (Placing the word in a chart) (P3-3)
-

In addition to the tools for explicit vocabulary assessment, the pre-service teachers will deploy the methods for implicit assessment and games. In other words, they will assess their future students' vocabulary knowledge and use through speaking and writing tasks. Of these, the speaking tasks will include role-plays, presentations, and talks about the content of audio or a text. As another way of producing, the writing tasks will consist of writing separate sentences, paragraphs, descriptions, answers to open-ended questions, preparing posters, or writing opinions after reading a text or listening to a recording. Moreover, P3-3 will prepare games through Wordwall and assess her students' vocabulary knowledge by asking them to match the words with the given pictures or place the correct word in the appropriate place in the chart.

Scoring, Interpreting, and Using the Assessment Results

Scoring the students' answers or performances is another essential step of the assessment process. Regarding this, the third-year PSETs mentioned the practices that were classified under the categories of either weak literacy or strong literacy. Teachers' practices of scoring according to their personal judgment or vague criteria constitute the category of weak literacy, while scoring according to a rubric, an answer key, or a set of specific criteria are the practices constituting the category of strong literacy.

Table 41

Scoring the Answers or Performances

WEAK LITERACY	STRONG LITERACY
<i>Scoring according to Personal Judgment</i> • Giving less points for the speaking assessment (P3-1) • Scoring according to errors (P3-1)	• Scoring according to a rubric (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6) • Scoring according to an answer key (P3-1, P3-2, P3-3, P3-4, P3-5, P3-6)
<i>Scoring according to Vague Criteria</i> • Scoring according to notes taken about the student's performance (P3-1, P3-4) • Not being sure how to score drawing after listening (P3-3) • Reading and listening text limiting the answers (P3-5)	• Inter-rater reliability (Another teacher rating the students' performance) (P3-5)

Regarding how they would score their students' performances, two PSETs gave answers indicating that they would score according to their personal judgment. In other words, they will not award scores according to a set of criteria they have determined in advance, but will make a rating based on their intuition. For example, P3-1 wrote in the scenarios that she would give less points for the speaking performances which would occur after listening to audio or watching a documentary about tourism rather than scoring according to any rubric or a set of criteria. Additionally, she wrote in the scenarios that when scoring the e-mails or letters of complaint written by adolescent students, she would either utilize a rubric or "or score each paper individually according to the rate of making mistakes without using rubrics." In the category of weak literacy, two participants also reported that they would give points according to the notes they took about the students' performances without using any rubric. Furthermore, while P3-3 stated that she does not know exactly how to score the drawings of the students, P3-5 pointed out that the answers to the questions about the reading and listening texts would be determined by the content of the listening and reading texts without any answer key.

In contrast to these weak literacy indicators, all the participants emphasized that they would use rubrics and answer keys in scoring, which are classified as strong literacy indicators. P3-5 also stated that, in addition to using rubrics, he would have another English teacher score his students' written and oral performances.

As for how the teacher candidates will interpret and use the assessment results, the findings disclose that they will carry out practices that can be classified under the category of strong literacy. These include the teacher trainees making interpretations about the effectiveness of teaching, adapting or changing the teaching methods, detecting the students' strengths and weaknesses, and making interpretations about the effectiveness of assessment methods. Additionally, the teacher candidates will provide feedback to their students by drawing on the assessment results and will determine whether their students are successful by comparing their scores with a cut-point score or making a norm-referenced interpretation. They will also focus on their prospective students' weaknesses to make improvements to the learners' language skills.

Table 42

Interpreting and Using the Assessment Results

STRONG LITERACY
• Making interpretations about the effectiveness of teaching (P3-1, P3-3, P3-5, P3-6) ○ Making an interpretation about the effectiveness of in-class activities (P3-1) ○ Questioning teaching methods if there are many students with 14 and below (P3-3) • Making an interpretation about the next teaching methods or activities ○ Making interpretations about the next listening activities (P3-1) ○ Making interpretations about the next speaking activities (P3-1) ○ Giving more writing assignments (P3-1) ○ Designing activities to focus more on writing skills (P3-1) ○ Individual interviews with the students and giving assignments to improve weaknesses (P3-2) ○ Making interpretations about how and what to teach (P3-5) ○ Using different methods if there is a problem with teaching (P3-5) ○ Making adjustments to teaching according to the results (P3-5) • Making an interpretation about the strengths and weaknesses ○ Monitoring the student's personal progress by comparing the previous results (P3-1) ○ Detecting the skills at which the students are competent or not (P3-3) ○ Detecting the reason for the students' incompetency in a skill (P3-3) ○ Making interpretations about the students' strengths and weaknesses (P3-5) ○ Detecting the weaknesses and trying to improve them (P3-5) ○ Detecting the students' strengths and weaknesses by comparing the CEFR (P3-5) ○ Making an interpretation about how well students improved their writing skills and language components (P3-5) • Making interpretations about the assessment methods ○ Questioning the effectiveness of assessment methods (P3-1) ○ Using the results while planning the next assessments (P3-1) • Teacher feedback (P3-4, P3-5) • Peer feedback (P3-4) • Making an interpretation about appropriateness of the curriculum to the students (P3-1) • Making interpretations about the students' learning strategies (P3-1) • Making a norm-referenced interpretation (P3-2) • Interpreting levels: Moderate-to be improved, poor-unsuccessful (P3-2) • Interpretation: Very good and good (Successful), Moderate (To be improved) (P3-2) • Detecting the skill that was less focused in the lesson (P3-3) • Making interpretation according to a cut-point score (P3-3, P3-4) • Making an interpretation: how the students can learn better (P3-5) • Making interpretation: how to adapt the course materials (P3-5) • Using the results to improve the students' speaking level (P3-5) • Detecting the weaknesses and trying to improve them (P3-6) • Detecting the most common errors (P3-6) • Determining the range between the highest and lowest score (P3-6)

Literacy in the Principles Component of LAL

Principles is the third component of LAL which refer to “ethics, fairness, democracy, and transparency” (Giraldo & Murcia, 2018). As compatible with this, the present research involves the findings classified under two main considerations in which the teacher candidates should be competent: having awareness of and ensuring ethical assessment process.

Table 43

Awareness of Ethical Assessment Process

STRONG LITERACY
<i>Unethical Assessment Experience</i> • Partiality among the students (P3-1, P3-4, P3-6) • Assessing what was not taught (P3-2, P3-4) • Inconsistency among the classrooms in terms of the midterm and final presentation dates (P3-2) • Ignoring the attendance (Just making your own presentation) (P3-3) • Assessing only grammar and vocabulary (Ignoring other skills) (P3-4) • The great difference among the classrooms in terms of the difficulty of the exams (P3-4) • Scoring the answers of the students from other classrooms (P3-4) • Not allowing the students to see their papers (P3-5) • Not explaining the reasons for the score (P3-6) <i>Features of the Ethical Assessment Process</i> • Impartiality (P3-1, P3-3, P3-5, P3-6) • Assessing what was taught (P3-2, P3-4, P3-5) • Avoiding sharing the questions before the assessment (P3-1) • Using inoffensive expressions in the feedback process (P3-2) • Informing the students about the assessment process with the syllabus (P3-2) • Checking by assessment what/how much was learned (P3-2) • Taking the students' characteristics and psychological backgrounds into account (As a "whole person") (P3-2) • Not disturbing the student or creating stress on him/her during the assessment (P3-2) • Blind assessment (concealing the students' names) (P3-2) • Scoring answers to the questions sequentially (P3-2) • Questions in the written exam not involving the answers (P3-3) • Not giving hints to some students during the speaking exam (P3-3) • Informing the students about the content and type of assessment at the beginning of the semester (P3-4) • Validity and Reliability of assessment (P3-4) • Using various assessment methods by considering the characteristics of different students (P3-4) • The assessment not being advantageous to a group of students (P3-4) • Scoring done by at least 2 raters (P3-5) • Teacher's having enough knowledge for quality assessment (P3-6) <i>The Effects of Parents, Other English Teachers, and Administrators on Fairness</i> • Parents' pressure on the teacher to raise the students' scores (P3-1, P3-3) • Other teacher's demotivating the teacher to conduct alternative assessments (P3-2) • Parents' criticisms about alternative assessment methods (P3-2) • Administration's demand from the teacher to share the questions with the students (P3-3) • Parents' pressure on the teacher or threats to the teacher (P3-4) • Parents' comparing and criticizing the students' scores (P3-6)

The findings reveal that the teacher candidates are aware of what an ethical assessment process is or is not and the potential effects of stakeholders on the fairness of the assessment. In order to disclose their awareness of what an unethical assessment is, they were asked to explain their assessment experiences, which they considered unethical. In this regard, the most frequently mentioned assessment experience is the teacher's partiality in favor of some students. Three of the participants referring to this experience emphasized that their teachers

gave higher scores to the students who could not have an adequately successful performance or did not take the necessary assessment. Furthermore, the teacher including the content or skills that s/he did not teach in the lesson into the assessment was an unethical assessment experience for two of the PSETs. Their explanation for this was that the content or skills not focused on in the lesson were advantageous for some students who had background knowledge regarding the content or were more competent in the target skills, while they were disadvantageous for other learners.

The features that the participants mentioned for an ethical assessment process were in line with the issues they stated regarding their unethical assessment experiences. That is to say, impartiality and assessment of what is taught are the most frequently cited features. In addition, they also referred to such features as students being informed about assessment in advance, impartial administration of the assessment techniques or tools, and reliable scoring of the students' answers or products.

The senior PSETs will also take part in a variety of activities to ensure ethical assessments. While some of these techniques may make it more difficult to ensure ethics in assessment, others are more successful at doing so. In this regard, the practices in the first group fall under the category of weak literacy, whereas those in the second group are classified as signs of high literacy.

Two of the indicators in the category of weak literacy include comparing a student's score with the class average rather than with a criterion score and giving all questions equal scores without taking into account the difficulty of the questions. In addition, not being aware of ethical problems that may be caused by factors such as gender, ethnic identity/race, and socio-economics and not being able to find solutions to these problems are also indicators of weak literacy. Other indicators of weak literacy also include sharing assessment results with students, parents, and administrators only through the electronic school system instead of sharing them openly and regularly or discussing them with these stakeholders upon their request. In addition, other components of the weak literacy category include not knowing how to ensure ethics in assessments to be conducted in online education or in assessments to be conducted with students from different nationalities.

In contrast to the indicators of weak literacy, planning to use appropriate methods to ensure impartiality, transparency, and democracy in evaluations shows that respondents have strong literacy. To exemplify, the teacher candidates will ensure that the teacher-student

relationship does not affect the assessment results and will take notes on the teaching and assessment process and use these notes in the assessment process. In addition, to ensure equality between classes, they will teach all classes equally and apply the same assessments to all classes. Moreover, to ensure transparency, they will give the exam papers to the students again after scoring and let them see their mistakes. In the preparation phase of the assessment, ethics will be ensured by preparing parallel forms, not including any content other than the taught content in the assessment, and choosing topics and tasks that appeal to students' common interests.



Table 44

Ensuring Ethical Assessment

WEAK LITERACY	STRONG LITERACY
<i>Ethical Assessment Process: Preparation, Administration, Scoring, and Interpretation</i>	<i>Solution for the Unethical Assessment Experience</i>
- Comparing the student's score with the mean (P3-1, P3-3) - Criterion not always giving the accurate information about the student (P3-1) - Assigning the same scores for all questions (by disregarding their level of difficulty) (P3-4)	- Taking notes (P3-2, P3-3) - No effect of the teacher-student relationship on the assessment results (P3-1) - Assigning scores for presentation, participation, and attendance (Dividing the scores into parts) (P3-3) - Using various assessment methods (P3-4) - Providing the same input to other classrooms and administering the same assessment to all classrooms (P3-4) - Giving the students their exam papers again (P3-5)
<i>Ensuring Fairness in Terms of Some Factors</i>	<i>Ethical Assessment Process: Preparation, Administration, Scoring, and Interpretation</i>
<i>Gender</i>	<i>Preparation</i>
- No effect of gender on assessment (P3-1) - Not exactly knowing how to ensure fairness (P3-6)	- Not asking the questions at a specific resource (e.g., a website) (P3-1, P3-4) - Asking about what was taught (P3-2, P3-6) - Preparing parallel forms (P3-5, P3-6) - Preparing the questions at a similar difficulty level (e.g., speaking exam with 3 questions) (P3-1) - Informing the students about how they will be assessed (P3-2) - Considering the content in the curriculum (P3-2) - Considering the consistency between the assessment methods suggested in the curriculum and the assessment tools prepared by the teacher (P3-2) - Choosing the topics by considering individual differences (Not using topics related to religion or politics) (P3-3) - Avoiding the behaviors that can give the students extra information about the exam (P3-4) - Being careful to use the tasks and give assignments appealing to all the students' interests (e.g., introverts, extroverts) (P3-4) - Giving group assignments for the introverts and considering the students' characteristics while deciding the groups (P3-4) - Determining the level of the students and preparing a test with regard to the level (P3-5) - Preparing the exam together with the board of English teachers (P3-5) - Preparing a tool compatible with the students' level (e.g., giving or not giving the words for a fill-in-the-blanks test) (P3-6)
<i>Ethnicity/Race</i>	
No idea about how to solve the problem (P3-2)	
<i>Socio-economic Level</i>	
- More related to macroeconomics (P3-2) - Not giving project assignments (P3-6)	
<i>Sharing the Assessment Results with Stakeholders</i>	
<i>Sharing the Results with Students</i>	
- Announcing the results in the classroom (P3-5) - Sharing the results on E-School (P3-5)	
<i>Sharing the Results with Parents</i>	
- Not sharing the results of the alternative assessments (P3-2) - No need to announce the results to the parents (P3-5) - Sharing the results if the student has serious problems (P3-6)	
<i>Sharing the Results with the Administration</i>	
- Not sharing the results with administrators (P3-1, P3-2, P3-5) - Sometimes, giving information about the type of assessment to the administration (P3-4) - Sharing the results if the administrators are interested (P3-6)	
<i>What to Do to Ensure Fairness in Online Education</i>	
Difficult to ensure fairness (many assignments and workload) (P3-4)	
<i>Ensuring Fairness in the Classroom Consisting of Students from Different Nations</i>	
- Not having enough knowledge (P3-1) - Not causing any unfairness (P3-3)	

-
- Trying to conduct a fair assessment (P3-6)

- Determining the criteria for different language components (e.g., x points for grammar, y points for vocabulary) (P3-6)

Administration

- Not helping the students during the assessment (P3-1, P3-2, P3-3)
- Recording the student's performance during the speaking assessment (P3-4, P3-5, P3-6)
- Preventing cheating during the exam (P3-1, P3-4)
- Students randomly selecting the topics (P3-5, P3-6)
- Arranging the seats (P3-5, P3-6)
- Disregarding the teacher-student relationship during the assessment (e.g., speaking exam) (P3-1)
- Allocating equal time for all students (P3-1)
- Giving roles/topics with equal difficulty levels (Speaking assessment) (P3-1)
- Being sure that the students understood everything related to the assessment (P3-2)
- Not distressing the students or empathizing with the students (P3-2)
- Listening to all students carefully during the presentations (P3-4)
- Providing feedback to all students equally (P3-4)
- The participation of two raters during the speaking assessment (Fair assessment) (P3-4)

Scoring

- Concealing the students' names (P3-2, P3-3, P3-5, P3-6)
- Scoring the answers sequentially (P3-2, P3-3, P3-5)
- Ensuring inter-rater reliability (P3-4, P3-5, P3-6)
- Deciding on the scoring before the assessment (P3-4)
- Using a detailed rubric (P3-4)
- Fair scoring without taking the students' characteristics into consideration (P3-6)

Interpretation

- Interpretation according to specific criteria (P3-4, P3-6)
- Paying attention to objectivity (Fairness) (P3-2)
- Interpretation according to the results of more than one assessment (P3-4)
- Giving feedback to the students (P3-6)
- Identifying the students' common weaknesses according to the assessment results (P3-6)

Ensuring Fairness in Terms of Some Factors

Gender

- Using the topics appealing to the common interests (P3-3, P3-4, P3-5)
- Ensuring an equal number of female or male students in the groups (for groupwork) (P3-4)

Ethnicity/Race

-
- Using more common topics in the assessment for the students of different races (P3-1, P3-3)
 - Choosing texts that are not offensive to students of any ethnicity or race (P3-4, P3-6)
 - Not giving extra points to the students from other races (P3-1)
 - Achieving a balance in focusing on cultures (P3-5)

General Knowledge

- Asking about the familiar topics or objects to the students (P3-1, P3-2, P3-3)
- Not asking about a specific culture (P3-1)
- Forming schemas about the topic in the students' minds with a presentation, photos, or videos if assessment is necessary (P3-1)
- Asking about the common world culture in culturally heterogeneous classrooms (P3-2)
- Expecting from the student what was taught (P3-6)

Socio-economic Level

- Asking the students about the necessary materials for assessment (P3-1)
- Using the topics familiar to the students (e.g., apple, orange instead of mango, avocado) (P3-3)
- No effect in the multiple-choice exams (P3-4)
- Being careful about the possible effects of socio-economic level in the task-based assessment (Flamboyant or simple assignments) (P3-4)
- Asking the students to buy economical materials (P3-5)

Ensuring Democracy in Language Assessment

- Students' voting for the topics given by the teacher (P3-1, P3-2)
- Getting feedback from the students about the assessment (P3-1)
- Teacher observing the students for detecting their interests (P3-3)
- Asking students about what they want to do (P3-4)

Ensuring Transparency in Language Assessment

- Sharing the checklist or rubric with the students before the assessment (P3-3, P3-4, P3-6)
 - Recording the students' speaking performance during the exam (P3-1)
 - Preparing criteria (P3-1)
 - Easy to ensure transparency for the multiple-choice exams (P3-1)
 - More difficult to ensure transparency for the open-ended questions (e.g., writing exams) (P3-1)
 - Informing the students about every step of the task by giving all instructions clearly and setting a time limitation (P3-2)
 - Recording what the students did in each assessment by considering the criteria (P3-4)
-

-
- Giving the students their papers and making explanations about the assessment results (P3-4)
 - Informing the students about the types of questions and the assessment procedures (P3-4)
 - Making all the objectives transparent (P3-5)
 - Assessing what was taught (P3-5)
 - Ensuring reliability (P3-5)
 - Informing the students about scoring after the assessment (P3-6)

Sharing the Assessment Results with Stakeholders

Sharing the Results with Students

- Sharing the results with the students individually (P3-1, P3-2, P3-4, P3-6)
- Sharing the assessment results with the group members in the group work assignments (P3-1)
- Giving the folders to the students to take a look at in the alternative assessments (P3-2)
- Sharing the score range in the classroom (P3-3)
- Sharing the frequent weaknesses or errors in the classroom (P3-3)
- Sharing all the criteria and papers with the students in groups (if sharing individually is not possible) (P3-4)
- Giving individual feedback to the students about their strengths and weaknesses (Feedback: Very important) (P3-4)
- Giving the exam papers back to the students (P3-5)

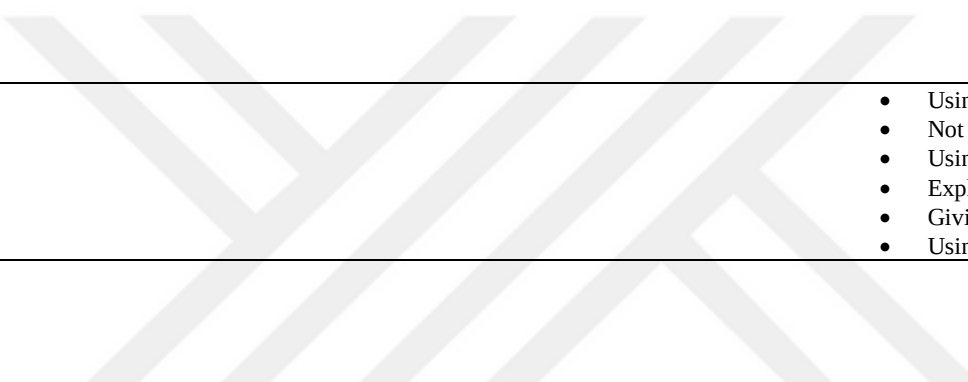
Sharing the Results with Parents

- Informing parents individually at parent-teacher meetings (P3-1, P3-2, P3-3, P3-4)
- Sharing the results on the phone by talking about the students' development (P3-1)

Ensuring Fairness for the Students with Low Vision

- Preparing or using assessment materials written in large fonts (P3-1, P3-4, P3-5)
- Using the Braille alphabet (P3-3, P3-6)
- Reading the exam aloud (P3-4, P3-5)
- Using alternative assessment methods (e.g., projects and performance assignments, video recordings) (P3-2)
- Using methods to express herself with movements (Listening assessment) (P3-3)
- Using methods that are easy to see and match (Listening assessment) (P3-3)
- Using embossed images on role cards (P3-3)
- Giving the student more time for assessment (P3-4)
- Allowing the student to use special pencils for writing (Easy to see) (P3-4)
- Using exam papers with thick lines and wide line spacing (P3-4)
- Getting help from special education teachers (How to assess and use materials) (P3-5)

Ensuring Fairness in the Classroom Consisting of Students from Different Nations

- 
-
- Using intercultural elements or content in assessments (P3-2, P3-4, P3-5)
 - Not using racial expressions in assessment tools (P3-2)
 - Using traditional and alternative assessment methods together (P3-2)
 - Explaining instructions with visuals and demonstration (P3-4)
 - Giving clear instructions and explanations in English (P3-4)
 - Using only English for assessments (P3-5)
-

In the administration phase of the assessment, the pre-service teachers will engage in practices such as not helping any student during the assessment, recording the speaking performances of the students, proctoring to prevent cheating, and ensuring that the students choose the speaking topics at random. In the scoring stage, they will perform practices such as scoring the exam papers by covering the names of the students, scoring the answers given by all students to each question, and moving on to the answers given to the next question. In addition, they will make interpretations by comparing the assessment results with specific criteria and providing feedback to the learners about their strengths and weaknesses.

In order to ensure fairness among students in terms of factors such as gender, ethnic identity or race, and general knowledge, they will pay attention to the selection of topics, the texts to be used as assessment material, and they will avoid practices that will cause unfairness among the learners. As for ensuring fairness in terms of socio-economic level, they will use materials that students can access easily and assessment tools that do not require students to spend money. Additionally, in order to ensure democracy, students' opinions about the evaluation methods will be elicited, and in order to ensure transparency, the PSETs will share the evaluation criteria with the students in advance and record their performances during the evaluation process.

Sharing the assessment results with the students and parents is another necessary step of the assessment process. To do this, the student teachers in Case 2 will hold individual conferences or have phone calls with the students and their parents and provide feedback regarding the learners' scores and their language learning performance. In these conversations, the pre-service teachers will also aim to find answers to the difficulties faced by the students in the language learning process.

Furthermore, they will also use large font assessment materials, the Braille alphabet, and alternative assessment methods for students with low vision. For students from different nationalities, they will make sure that the topics used are not offensive, that they are about a culture that all students are familiar with, and that instructions for assessment are given in English only.

Reflections, Suggestions, and Expectations for the Improvement of LAL

This section comprises reflections on the LAL training they received, expectations from the TEELT course, and suggestions regarding this course. In this vein, the findings reveal that

P3-2 and P3-5 mentioned their experiences in the ELT department which they found beneficial for the improvement of LAL. One of these was the assessments used by the two instructors whose courses P3-2 took in the English language teaching program. P3-2 emphasized that these assessments were an example for him and that he took them as role models. P3-5 also agreed with P3-2 and added that he found the feedback sessions within the Teaching Language Skills helpful.

Table 45

Reflections, Suggestions, and Expectations for the Improvement of LAL

Reflections Regarding the LAL Training <i>Reflections on LAL Development in General</i> • The very positive effects of two professors on LAL (role models) (P3-2, P3-5) • Positive effects of feedback sessions in the Teaching Language Skills course on LAL (P3-5) <i>Reflections on the MEE course</i> <i>Negative Reflections</i> • Negative effects of taking MEE course online (P3-2, P3-4, P3-5) • Not knowing how to use the theoretical knowledge learned in the MEE course (P3-3) • Allotting too little time for the exam (10 minutes) (P3-3) • No explanation about the reason for the assessment (P3-4) • Not getting any feedback about the assessment results (P3-4) • The MEE course not appealing to the ELT program (P3-4) • No significant contribution from the MEE course (P3-4) • Students from other departments taking the MEE course together (P3-5) <i>Positive Reflections</i> • Getting theoretical knowledge from the MEE course (P3-3, P3-5) • Learning how to calculate discrimination and difficulty indexes (P3-2) • Learning some terms and calculations in the MEE course (P3-4) Expectations and Suggestions Regarding the LAL Training <i>Expectations for the LAL Development</i> • Professors in the ELT program using alternative assessment (Better) (P3-3, P3-5) • Getting practical knowledge from the TEELT course (P3-5) <i>Suggestions for the TEELT course</i> • Integrating the course with the practicum for practice (P3-2, P3-4, P3-5) • Focusing on how to use assessment results (P3-1) • Focusing on how to solve the problems faced in the assessment (P3-1) • Focusing on how to assess students at different levels (P3-2) • Focusing on how to assess refugee students (P3-3) • Focusing on how to assess students with special needs (P3-3) • Focusing on validity and reliability (P3-3) • Focusing on how to ensure fairness (P3-3) • Preparing exams in different formats (P3-4) • Getting the professor's and peers' feedback about the exams (P3-4) • Integrating the course with the Integrated Language Skills course (P3-5)

The participants also reflected on the MEE course. Regarding this course, three of them emphasized that the fact that the course was online caused them not to benefit from the

course. However, However, two PSETs emphasized that this course only helped them to learn information that would be useful for the Public Personnel Selection Examination.

As for their expectations for the LAL development, P3-3 and P3-5 also expressed their expectations for LAL development and said that they expect teacher trainers in the ELT program to use alternative assessment methods. Moreover, P3-5 stated that he would like to practice as well as have theoretical knowledge in the TEELT course.

Participants in Case 2 made various suggestions for the TEELT course they will take next year. Participants in Case 2 made various suggestions for the TEELT course they will take next year. For example, they suggested that the TEELT course and the teaching practice course should be coordinated. They also suggested that this course should focus on the assessment of students at different levels, refugees or students with special needs.

Case 3: Fourth Graders

This case consists of five senior English teacher candidates. The courses they take in the first semester are “Atatürk's Principles and History of Turkish Revolution I”, “Materials Adaptation and Development in Foreign Language Teaching”, “School Experience”, “Guidance”, “Special Education”, “Second Foreign Language III”, “Teaching Integrated Language Skills”, and “English in Mass Communication (elective)” (Gazi University, 2020). In the second semester, they take the courses “Atatürk's Principles and History of the Turkish Revolution II”, “Testing and Evaluation in English Language Teaching”, “Comparative Education”, “Turkish Education System and School Management”, “Teaching Practice”, “Semantics and Language Teaching (elective)”, and “Sociolinguistics and Language Teaching” (Gazi University, 2021). Of these courses, Testing and Evaluation in English Language Teaching (TEELT) is the main and compulsory course that all English teacher trainees take in the last semester of the ELT program. The TEELT aims to provide PSETs with the basic knowledge of foreign language assessment and to enable them to practice the assessment methods and tools that they can use to assess language skills.

The first week of the TEELT course focuses on the historical development of testing, the importance of testing, the concepts of testing and evaluation, and the classification of tests. In the following weeks, how language components could be assessed were covered. Of these, two weeks were devoted to vocabulary tests and one week to grammar and pronunciation

tests. In these lessons, limited response items were first introduced to the PSETs. Later, more open-ended methods for assessment were mentioned. After covering these subjects, the instructor of the course assigned the teacher candidates to prepare tests as midterm exams. Within the scope of these assignments, the PSETs prepared limited and free response items for each language component. For example, they prepared multiple-choice questions and fill-in-the-blank items for assessing vocabulary.

MULTIPLE CHOICE COMPLETION

Book: Interchange Level 3

Unit: 5th unit / Crossing cultures

1. The boy's mother felt because he didn't respect his friends' ideas.

A-) Happy B-) Confident **C-) Embarrassed** D-) Excited

Figure 5. Multiple-choice question for vocabulary assessment (Prepared by P4-1)

However, in addition to these items, they also prepared a cloze test and items that required students to listen to commands and write a word or perform a behavior, and items that required students to listen to commands or questions and respond by pointing to a picture or saying "yes" and "no".

Look at the picture. Read the questions below and point to the jobs. (Face2Face: Elementary. Unit : 7/B Jobs Page: 43-44)



1. Who is wearing yellow?
2. Who are wearing white?
3. Who is wearing blue?
4. Who is wearing a hat?
5. Who is wearing an astronaut sweats?

Read the commands below and start drawing.

Draw a vet.
 Draw a singer.
 Draw a painter.
 Draw a dancer.
 Draw a pilot.

Figure 6. Limited response item for vocabulary assessment (Prepared by P4-2)

After the midterm exam, the focus is on the assessment of language skills. In this context, limited response tests as well as more open-ended items such as reading and paraphrasing are mentioned in relation to the assessment of reading skills. The following week is devoted to the assessment of writing skills. In this course, there is a progression from more restricted response items such as sentence combining or sentence expanding to more free response assessments. The evaluation of listening skills is the topic of the next week. Evaluation of listening skills is the topic of the next week. In this course, as in other courses, there is a course flow starting from limited response items to assessment methods in which students can express their thoughts verbally or in writing after listening to. In the course devoted to the evaluation of speaking skills, the course content started from limited response items (such as directed responses and questions about pictures) and progressed to more free response methods such as guided techniques and interviews. In the last week, concepts such as reliability, validity, item discrimination and item difficulty were covered within the scope of the evaluation of the tests.

Teacher candidates again prepared test items as their final homework. These items were prepared in line with the course process, starting from limited response items to free response items. However, it should be noted here that the participants only prepared items to evaluate reading and writing skills. For example, they prepared true-false statements for assessing reading skills as in the figure below.

B. SENTENCE COMPREHENSION
(New English File: Elementary, I can't dance p 40-42)

TASK NUMBER 2:

2.



a) Look at the picture above and read the sentences. Circle T if you think the statement is true or circle F if you think the statement is false.

The boy with the sunglasses is sunbathing. (T/F)

A boy and a girl are playing table tennis. (T/F)

The man with a hat is carrying watermelons. (T/F)

Answer key: 1. T, 2. F, 3. T.

Figure 7. Limited response item for vocabulary assessment (Prepared by P4-3)

In addition to these limited response items, they also prepared free response items. To illustrate, P4-4 prepared a writing task requiring the students to write a story based on the given pictures.

Free Writing

1. Level: Intermediate

The book's name: Total English

Unit: 7.1

Page: 92

- Write a short story by looking at the pictures.



Figure 8. Free writing task (Prepared by P4-4)

As for the participants, P4-1 is a 22-year-old male teacher candidate. In his first opinion journal, he wrote that from the fourth grade of the primary school, he started learning English, to the tertiary level, he was assessed with multiple-choice and fill-in-the-blank questions. Therefore, he stated that he did not remember answering any open-ended questions until university, and he wrote that “my opinion wasn’t taken into account or nobody assessed my language proficiency except for reading”. In fact, this ignorance of other skills was at such a level that he expressed this situation with this sentence: “When it comes to the other three main skills and sub-skills of English, I wasn’t aware that learning a language wasn’t only about being able to understand what you read”. Moreover, he adds that all the activities he did in English lessons were restricted activities, not requiring him to use the target language communicatively. Interestingly, he also reported that he had not been taught vocabulary in English lessons and therefore learned new words on his own through constant repetition.

P4-1 would not have forced his learners to memorize grammar rules and long lists of words. In addition, he would have raised the students’ awareness about the language learning process and all language skills. Regarding this, he noted that he felt so unconfident when he was asked to use his language skills communicatively at the university. Therefore, he would have enabled his students to improve their English by collaborating with their peers and fulfilling the given tasks. As another practice that he would not ignore, he emphasized that he would have gotten the learners’ opinions through the open-ended questions and tracked their language development by examining their answers. The open-ended questions would also enable the students to be more autonomous and active during the language learning process. He would also have frequently made in-class observations and assigned scores by taking their participation into account. As the reason for this, he explained that “each learner doesn’t have the same learning style, and they are, individually, too different to be labeled after an exam. They couldn’t reflect their knowledge on the exam paper, but they might express what they know in classroom activities”. In addition to these observations, he would have assigned some homework since the internalization of language could not be confined to only the classroom. He would also have “[tried] to initiate an authentic language learning atmosphere rather than compelling learners to pass the exam”.

P4-2 is a 33-year-old Turkmen female pre-service teacher. She started learning English in kindergarten and learned the alphabet and some very basic words, like the names of animals.

She had the same English teacher from primary school to high school. Although she could not remember exactly what they did for the assessment, she stated that they memorized poems to improve their speaking skills and pronunciation. Moreover, her teacher acted out stories with puppets. Then, while she was studying in high school, she also acted out some stories with puppets as a part of the speaking assessments. She had native English-speaking teachers, and she had to talk to them for the assessment of her speaking skills. Furthermore, she was asked to comment on a given picture in these assessments. In the reading assessments, she answered open-ended questions, and she emphasized that multiple-choice questions had almost never been used in assessments. As an example of how the open-ended questions were used, she stated that she was given a text and asked “whether she would like to go to that city” and “why she would (not) like to go there”. However, listening skills were often neglected, although they sometimes listened to some songs in lessons. Therefore, she said it was not asked of her to listen to a recording and fill in the blanks or interpret what he understood. As the reason for this, she expressed that there were not adequate facilities to listen to the recordings. For the assessment of writing skills, she stated that dictations were done and three or four different topics were given to her, and she was asked to write a text by choosing one of them. In addition to this, she had to memorize a certain number of words each week.

P4-2, having experienced more communicative assessment methods, would have conducted similar assessments. For example, she wrote that she would also have asked open-ended questions, given reinforcement and positive feedback, and assessed her students’ language development in the process. However, she would have focused a bit more on grammar since she sometimes had problems with it. Also, she would have avoided multiple-choice questions as much as possible but employed them to test her knowledge of vocabulary. Observations would also have been important for her assessments, and she would have given feedback to her students according to these observations. To exemplify, she would have provided feedback immediately after the learners made a presentation because it would have made them realize their strengths and weaknesses and improve themselves.

P4-3 is a 21-year-old female teacher candidate. Although she had to start learning English in the 4th grade, she literally started taking English lessons in the 6th grade because there was no English teacher. “[Her] primary school teacher tried to take that job, but after half a semester, s/he said s/he couldn't do it and started teaching math in these lessons.” The

assessment methods or techniques she experienced were similar in lower and higher secondary schools. The assessments she took consisted only of written exams. “These exams usually start with a reading text, followed by classical open-ended questions, fill-in-the-blanks, and finally word matching. There were also multiple-choice questions, but these were much rarer”. She stated that “[her teachers] told them that they did not want to ask multiple-choice questions in order to prevent cheating and/or not to simplify the questions”. In addition to these items, they were very rarely expected to order the mixed parts of a reading text in order to create a meaningful whole. Since listening and speaking skills were neglected, she commented that these assessments were not effective enough. She also added that her teachers asked very easy questions in the exams so that their students would get high scores. Moreover, she commented that she did not prefer the fill-in-the-blank items because they forced the students to memorize if the subject of the assessment was knowledge of grammar. In contrast to this, she expressed that ordering the mixed parts of a given text was a good assessment technique because it differentiated the students who were more proficient in English from those who were less proficient. She also got scores from performance assessments, which involved writing a composition and making a presentation. Regarding the scoring of the students’ presentation performances, she thought that her teachers were not impartial.

If she were her English teacher in lower and upper secondary school, she would also organize a test of listening skills since the schools had the necessary system to do that. She would have organized a speaking exam, for which discussion groups could have been arranged, or two students could have been expected to talk to each other. In the questions about matching words with their meanings, I would set the words not with their Turkish meanings but with a picture or with their definitions in English. Students are already trying to learn English entirely through translation, which is quite wrong, but teachers almost encourage this by making Turkish-English pairings. In written exams, I could use fill-in-the-blanks to measure vocabulary knowledge, but I would prefer to include dialogue completion questions to measure grammar knowledge. Dialogue can provide more background for students, and I think it would be good for them to be able to follow a dialogue in general. I like the questions about sorting the mixed reading text into a meaningful whole. I would apply this in the same way. It distinguishes the students who know the language very well from the students who

do not. For higher-level students, I could ask them to write an essay or paragraph at the end of the exam.

P4-4 is a 21-year-old female teacher candidate. In the 4th grade, she started learning English. Although she had many English teachers from primary school to upper secondary school, most of them used visuals and videos to make their students like English. Some of these teachers awarded their students with small gifts like chocolate when they gave correct answers and participated in the lesson, and she remembered that she ate a lot of chocolate thanks to this type of feedback. In primary and lower secondary school, she took written exams involving multiple-choice and open-ended questions. Although listening, reading, and writing skills were focused on in the lessons, reading and writing skills were measured through the tests. According to P4-4, this is the wrong assessment practice.

Throughout her training in upper secondary school, she had three English teachers. Until becoming a language major, she took tests similar to those in primary and lower secondary school. In the 11th and 12th grades, she was a language major, and her teacher gave daily assignments and asked the students to answer 250 multiple-choice questions every day. Moreover, the teacher scored these assignments and the questions. However, P4-4 explained in the interview that the daily assignments involving the fill-in-the-blank items on an exercise book were given to the students with low English levels. Therefore, she stated that she had never been given these assignments. Also, in line with the university entrance exam, she took mock exams every week, and her teacher took their results into account for the assessment.

All language skills would be assessed through various assessment methods rather than only open-ended and multiple-choice questions. To clarify, she emphasized that the assessments would have integrated the language skills (although some assessment practices she mentioned were not like that). For example, she would have assessed her students' speaking skills through role-plays and given them scores by observing them. They would have also written a text or an essay about the topic chosen. In the listening assessments, they would have filled in the blanks and answered open-ended questions by listening to audio twice. The reading assessments would have involved multiple-choice questions asking the learners to find the best title and open-ended questions requiring the learners to make inferences about the events mentioned in the text. Students would have been awarded prizes for their

achievements. However, it will take care not to cause jealousies to arise by triggering feelings of competition among students. In upper secondary school, since the class was not crowded, she would have preferred to give individual assignments. These assignments would have consisted of answering multiple-choice questions.

P4-5 is a 22-year-old female pre-service teacher. Although she could not remember it clearly, she stated that there were sections on writing the Turkish equivalents of words and sentence completion in the exam paper. She pointed out that she liked these exams because she liked memorizing words and writing them 10 times. In the 7th grade, she had an English teacher who she liked very much and encouraged her to learn English. The lessons taught by this teacher were very fun, and the students mostly participated. Also, she wrote that the teacher did an oral exam on words every week and included them in her assessments. In the 8th grade, she had another teacher who encouraged the students to speak English in the first lesson. But in the following lessons, she did not speak English herself and did not encourage the students to do so either. The assessments done by this teacher were written exams in which fill-in-the-blanks and writing the Turkish or English equivalents of the given words were used.

At the beginning of upper secondary school, she was administered a placement test. Although P4-5 could not remember exactly, she wrote: "There was a wide variety of questions. I don't remember exactly, though. At the end, there was a 30-point essay". Then the exams administered until the 12th grade involved multiple-choice questions and fill-in-the-blanks. In addition to her grammar and vocabulary knowledge, her listening skills were also measured through these exams. In the 12th grade, an oral exam was conducted to assess speaking skills. However, this was a 5-minute exam in which the students were expected to answer specific questions asking about their names, ages, etc.

If she was in her teacher's place in primary school, since they loved singing songs at that time, she was sure that exams with songs, singing that song, and asking questions about it would have been very enjoyable and fun for all of them. Fill-in-the-blank items would also have been employed, and she would have asked the learners to sing that song. Also, she would ask inference questions, requiring the learners to provide answers regarding the events, characters, and places. Short stories could also be included, and the students would have acted them out. Regarding this, she remembered the theater play they performed in Turkish class when she was in the second grade of primary school. Moreover, the students

would have written short stories using the words given by the teacher. In the speaking assessments, the learners would have drawn a topic randomly out of a box and given a short talk about it. In the 8th grade, she would have focused more on the topics appealing to the students' interests, utilized enjoyable activities, and spoken in English during the lessons. In this way, the students would have played roles or performed scenes from a movie or a theater play. In fact, she would have adopted stories, role-plays, and individual talks for assessing the students at all levels of English.

Literacy in the Knowledge Component of LAL

Knowledge about the Key Concepts of Assessment

In this section, the findings regarding the senior teacher trainees' knowledge about the key terms specific to educational and language assessment are presented. As in other sections, the findings obtained from the data collected through the first and third semi-structured interviews are presented separately in order to detect the effects of the TEELT course on the participants' knowledge of the key concepts. In this regard, the findings from the first interview reveal that most participants in Case 3 are not knowledgeable enough about some key terms such as test, assessment, evaluation, reliability, validity, and washback. In contrast to this, few participants have the necessary knowledge about such concepts as test, assessment, measurement, evaluation, validity, practicality, and washback.

According to the findings from the third semi-structured interview, most of the senior pre-service teachers are not knowledgeable enough about the concepts of assessment, measurement, evaluation, validity, and washback, while some of them have sufficient knowledge to explain or define these key terms. However, as the findings imply, the teacher candidates generally could not increase their knowledge of the basic terms of assessment, although they took the TEELT course in which this content was covered.

Table 46

Knowledge about the Key Concepts of Assessment

Categories	Codes (Semi-structured Interview I)	Codes (Semi-structured Interview III)
WEAK LITERACY	<i>Test</i>	<i>Test</i>
	Multiple-choice questions (P4-1, P4-2, P4-3, P4-4, P4-5)	Multiple-choice questions (P4-3, P4-5)
	<i>Assessment</i>	<i>Assessment</i>
	- Necessary to focus on more than one structure or skill (P4-2) - Assessing all the content taught for a year (P4-2) - Formal (P4-2) - Measuring the students' readiness levels (P4-3) - Making interpretations based on data like the student's level, abilities, or vocabulary knowledge (P4-3) - Making detailed comments on what the students know and don't know at the end of the test (P4-5)	- Involving scoring (P4-2, P4-4) - Done at the end of the semester (P4-2, P4-5) - An assessment administered to revise and detect the students' level by focusing on more than one topic (P4-1) - Measuring how much is taught and learned (P4-2) - A broader term involving reliability and validity (P4-2) - The step after testing (P4-2) - A more comprehensive term (P4-3) - Involving making a decision (P4-4)
	<i>Measurement</i>	<i>Measurement</i>
	- The same with assessment (P4-2) - Used to decide when students will move from one level to another (P4-2) - More concrete (P4-4) - Quantification of something like measuring length or degree (P4-4) - Unable to define for educational assessment (P4-4)	- Measuring the whole (P4-2) - Not involving scores (P4-2)
	<i>Evaluation</i>	<i>Evaluation</i>
	- Involving testing and assessment (P4-1) - Peer feedback for a pre-service teacher (P4-2) - In-service teachers' feedback to another teacher about his or her teaching (P4-2) - Observing the students' social relationships (P4-4) - A comprehensive comment about the student's performance at the end of the semester (P4-5)	- An umbrella term involving scoring and analyzing the students' performance or the knowledge they gave on their exam papers (P4-1) - Not involving scores (P4-2) - Involving a comparison of the performance with the criteria (P4-2)
	<i>Reliability</i>	<i>Reliability</i>
	Item difficulty index (P4-1)	- Accurate scoring (P4-3) - Distinguishing the students who know from the ones who don't (P4-3) - Separating the right answer from the wrong one (P4-5)
		<i>Intra-Rater Reliability</i>
		Unable to define (P4-2, P4-5)
		<i>Inter-Rater Reliability</i>
		Unable to define (P4-2, P4-5)
		<i>Methods to Estimate the Reliability</i>
		Not having enough knowledge (P4-1, P4-2, P4-4, P4-5)

-
- The stem and the options not involving any errors (P4-2)
 - The existence of strong distractors (P4-2)
 - The questions not being too easy (P4-2)
 - Test questions distinguishing the students who know from those who do not (P4-3)
 - Test questions not varying from person to person (P4-3)
 - Stating an example given by the instructor of the MEE course (e.g., Research on sleep, but a person acts like sleeping) (P4-4)
 - Unable to define (P4-5)

Intra-Rater Reliability

- Unable to define (P4-1, P4-2, P4-5)
- The rater recalculating the scores (P4-3)
- The rater checking whether s/he scored accurately (P4-3)
- Trust in the rater (P4-4)

Inter-Rater Reliability

- Unable to define (P4-2, P4-4, P4-5)
- Comparison among the teachers administering the same or different assessments (P4-1)

Methods to Estimate the Reliability

- Not having enough knowledge (P4-1, P4-2, P4-3, P4-4, P4-5)
- Having calculation experience with Excel (P4-3)

Validity

- Fairness among the students (i.e., assessing the students by considering their capabilities) (P4-1)
- Unable to define (P4-4)

Types of Validity

- Not having enough knowledge (P4-1, P4-2, P4-3, P4-4, P4-5)

Practicality

- Assessments enabling the teacher to draw conclusions about the students' weaknesses (P4-4)

Washback Effect

- Proactive inhibition (P4-1)
- Retroactive inhibition (P4-1)
- Stopping learning as a result of external factors (P4-1)

Validity

- Not having enough knowledge (P4-3, P4-4)
- An assessment based on some specific, predetermined criteria (P4-1)
- Too strong distractors reducing validity (P4-2)
- Deleting every 5th, 7th, or 10th word in the Cloze test (P4-2)
- Scores repeating more than once (P4-3)

Types of Validity

- Not having enough knowledge (P4-2, P4-3, P4-4)

Washback Effect

- Unable to define (P4-1, P4-2)
 - The demotivating effect of exams on student's learning (P4-3)
 - Errors becoming permanent because of the lack of reassessment (P4-5)
-

	• Unable to define (P4-2) • The stressing effect of the upcoming exam on the student and lowering the student's success (P4-3) • The exam causing the student to forget what s/he has memorized afterwards (P4-5)	
STRONG LITERACY	<i>Assessment</i> • Involving the exams and participation (P4-1) • Involving various types of questions like open-ended questions (P4-1) • Involving all assessment methods and tools (e.g., role-play) (P4-4) <i>Measurement</i> • The score given at the end of an exam (P4-1, P4-3, P4-5) <i>Evaluation</i> • Making an interpretation about the student's level and success by comparing the score with a cut-point score or criterion (P4-3) <i>Inter-Rater Reliability</i> • The consistency between the scores given by at least two raters (e.g., 100 versus 50 points given to the same test) (P4-3) <i>Validity</i> • Asking about what was taught (P4-2, P4-5) • The exam measuring the student's actual success (P4-3) • Limited validity (the student gets 100 points from the exam without having enough competence in English) (P4-3) <i>Practicality</i> • Economical (P4-1, P4-3, P4-5) • Feasible (P4-1) • Multiple-choice questions (Very practical; Easy to score and appropriate to use in different fields) (P4-2) • Open-ended questions (Not practical for scoring) (P4-2) • Easy to answer (P4-2) • Easy to access and use (P4-3)	<i>Test</i> • Assessment tool (P4-1, P4-2, P4-4) <i>Assessment</i> • Collecting information about the student's performance in language skills (e.g., speaking) (P4-3) <i>Measurement</i> • Scoring system used after the assessment process (P4-1) • Denoting something with numbers and symbols (P4-3) • Not making an interpretation (P4-4) • Obtaining a numerical value (P4-4) • Reaching a measure according to the rubric (P4-5) <i>Evaluation</i> • Coming to a decision by comparing the measure with a criterion (P4-3) • Detecting the students' weaknesses (P4-4) • Deciding which skills to improve (P4-4) • Coming to a judgment as a result of all assessments (P4-5) <i>Reliability</i> • Consistency between the scores after a test is administered twice (P4-2, P4-4) • Taking a value between 0-1 (The value of correlation) (P4-1) <i>Intra-Rater Reliability</i> • The same person scoring the same product twice within a while (P4-1, P4-3, P4-4) <i>Inter-Rater Reliability</i> • Consistency between the scores given by different raters (P4-3, P4-4) • Emerging as a result of two people evaluating a situation (P4-1) • Correlation between the raters (P4-1) <i>Methods to Estimate the Reliability</i> • Cronbach's alpha (P4-3) • KR-21 (P4-3)

• Easy to score (P4-5) • Time-saving (P4-5)	<i>Validity</i> • Content validity (P4-1) • Predictive validity (P4-1) • Asking the student about what was taught (P4-5)
<i>Washback Effect</i> • Students studying in line with the exam content (P4-3) • The effects of the high-stakes exams on the students' language learning (e.g., studying the language skills and components focused on in the high-stakes exam) (P4-4)	<i>Types of Validity</i> • Content validity (P4-1, P4-5) • Predictive validity (P4-1, P4-5) • Face validity (P4-5) • Construct validity (P4-5) <i>Practicality</i> • Economical (P4-1, P4-2, P4-3, P4-4, P4-5) • Easy to use without much effort (P4-1, P4-2, P4-5) • Suitable for use with the available materials (P4-3) • Easy to prepare (P4-4) • Appropriate to use more than once (P4-4) • Easy to access the assessment tool (P4-5) <i>Washback Effect</i> • Effects of high-stakes exams on classroom-based assessment (P4-3) • Positive washback effect (P4-3) • Negative washback effect (P4-3) • Teaching according to the content of the exam (P4-4)

Weak Literacy of the Key Concepts of Assessment (First Semi-structured Interview)

In the first interview, all of the teacher trainees defined the test as multiple-choice questions, while P4-1 added that the test also involves other limited response or “mechanical” items. As a remarkable answer, P4-2 stated that “tests serve to determine how much a person can memorize what s/he has learned without confusing it with other information” and made the following interesting analogy to explain the multiple-choice questions that make up the content of the test:

There is something like this in the test: Let's think of a machine. Let's understand that machine as the question of the test, and we leave something missing in the machine, and there are a few tools next to it. Which one of them should he use to complete that deficiency? Will he choose the right option? This is what a test looks like to me (Semi-structured Interview I).

Assessment was one of the concepts about which three participants in Case 3 did not have accurate knowledge. According to P4-2, assessment means the formal procedures focusing on more than one subject and aiming to detect the learners’ knowledge. Moreover, she adds that the assessment is conducted to measure how much the students have learned about the subjects covered in one year. P4-3 expressed that she did not have detailed knowledge of the concept and made the following description of the assessment: “It is something that serves to measure the current level of readiness of the students and to see what the students already know”. In line with this comment, she made the following statement, indicating that she confused the concepts of assessment and evaluation: “Assessment is a bit more like making comments based on mathematical data. For example, [P4-3]'s level may be C1. She can make these comments; she can speak in this way; she knows these words”. Similarly, the definition made by P4-5 was that assessment is to comment on students' strengths and weaknesses after the testing process:

Assessment, in my opinion, is the teacher's evaluation of the result of the test, like a form, like a form with five options. In other words, the teacher reaches a conclusion based on the data obtained as a result of the test. I think it is a situation where the teacher can say that “s/he knows this, but s/he may not know that, or s/he does not know this, but s/he is good at these” (Semi-structured Interview I).

As for the measurement, two participants, P4-2 and P4-4, provided definitions incompatible with the actual meaning of this concept. When asked, P4-2 first said, “I actually feel like I'm seeing it for the first time”. However, she later reported that assessment and measurement are the same and commented that measurement can be used to decide whether students will move from one language level to another. On the other hand, P4-4 said that she could not

define this concept in terms of assessment, although the concept of measurement reminded her of measuring the length, width, or degree of something.

Evaluation, which means the procedure of making a decision by comparing the data collected through assessment or testing with a set of criteria (Bachman, 1990, pp. 22-23), was incorrectly defined by four teacher candidates. According to P4-1, evaluation is an umbrella term that includes all other concepts like test, assessment, and measurement. As different from this, P4-2 based her comment on her experiences with micro-teaching and defined the feedback given by peers to a pre-service teacher or the feedback of in-service English teachers to another in-service English teacher about his or her teaching performance as examples of evaluation. P4-4, on the other hand, considered this concept from a more social perspective and defined it as "observing students' social relationships at school". According to P4-5, evaluation is the teacher's general comment about what the student can and cannot do at the end of the semester.

Reliability is the concept that all senior pre-service teachers in the research had problems defining accurately. In their definitions, these teacher candidates confused reliability with other concepts such as item difficulty index, item difficulty, item discrimination, or validity. Moreover, P4-5 wanted to skip the question by saying she would not be able to define the concept after she hesitated between not asking the wrong question and not contradicting the question with another one in the test.

As the concepts related to reliability, intra- and inter-rater reliability were also asked of the participants. Regarding the first one, none of them succeeded in defining or explaining it, and three of them could not even give any interpretation. Of the remaining two participants, P4-3 explained it as a recalculation of the test results, as in the excerpt below, while P4-4 guessed that it might be the trust of the assesseees in the assessor that s/he would not discriminate among his or her students.

Intra-rater reliability. Let's just say I'm the assessor. After I calculate the test results, let's say someone else calculates them, or I can find out whether I have calculated them correctly or not. You know, when we do a test in mathematics, we verify that we have reached the result correctly. I think of it as something like that (Semi-structured Interview I).

As for the inter-rater reliability, three participants were unable to provide any answers to define it. On the other hand, P4-1 reported that inter-rater reliability means the calculation of the mean of the scores awarded by different teachers administering the same assessment or different assessments. Furthermore, none of the teacher candidates in Case 3 remembered

any of the methods to estimate reliability. Regarding this, P4-3 stated that she could not remember them, although they had made some calculations through Excel in the TEE course.

Validity, which is another very important principle of assessment (Brown, 2004, p. 22) and means the appropriateness, meaningfulness, and usefulness of the interpretations made from the assessment results for the purpose of assessment (Gronlund, 1998, p. 226), is another concept about which two senior PSETs were not knowledgeable. While one of these PSETs, P4-4, could not answer the question at all, the other, P4-1, gave an answer unrelated to the concept. In his comment, P4-1 made a connection between the validity and fairness of the assessment and explained it by referring to the following example that the TEE course's professor provided:

There were three different animals, if I remember correctly. There was a question like elephant, monkey, and bird, and the one who climbs the tree wins. Of course, the most advantageous ones were the monkey and the bird. But the elephant is not evaluated in the same way. For this reason, I remember that it remained low in the validity part. I think validity may be evaluated in terms of equality, that is, fairness (Semi-structured Interview I).

The participants were also not familiar with the types of validity or types of validity evidence, and therefore no pre-service teacher could answer the relevant question in the interview.

The final two key concepts for participants to define were practicality and the washback effect. All but one of the participants were familiar with the former, while the other three were not knowledgeable about the latter. That is to say, P4-4 commented that if the results of an assessment enable her to make interpretations about her students' weaknesses, the assessment can be called practical. As for the washback effect, P4-2 could not comment on it, while the others referred to inhibitions or influences of the exam on the students' memory. Of these, proactive and retroactive inhibition were mentioned by P4-1 because the washback effect reminded him of the concept of inhibition taught in one of the educational sciences courses he had taken in the first grade. Unlike him, P4-5 gave such a definition of the washback effect:

You give a test to see if the student knows or doesn't know. But when you do this, it has a bad effect on the student. We forget it completely after memorizing it. I think this is a washback effect. [...] Because you study completely for two days. Then, when I leave that exam, or it can even happen in the exam, they are completely forgotten. What does this do? It pulls you back more than you learn. That's why the washback effect emerges (Semi-structured Interview I).

Strong Literacy of the Key Concepts of Assessment (First Semi-structured Interview)

Some of the senior teacher candidates could also define or explain some key concepts, and these were categorized under the category of strong literacy. Of these, two participants provided accurate explanations of the assessment concept. Regarding this, P4-1 reported that assessment is a broad term that involves assessment methods other than written exams. In line with this, he emphasized that the teacher's observations of the students' participation in the lesson and various questions, such as open-ended questions asking the learners to express their opinions, can be employed in the assessments. P4-4 also pointed out that assessment is a broad concept and can include many different methods or techniques. As examples of these, she cited role-playing and the teacher's observations made for assessing the students' speaking skills.

For defining measurement, all four teacher trainees referred to the quantification of the assessment results. P4-1 defines the measurement as "the grade obtained as a result of an exam." Likewise, P4-5 said that scores out of 100 and grades like AA and BA based on a numerical scale are called measurements. According to P4-3, measurement is "something that is generally a bit more formal, based on more rigid things, just showing a score, and so on."

P4-3 was also the only participant who could accurately define evaluation. The definition given by her was "the process of arriving at a conclusion based on certain criteria" from the scores or assessment results obtained through testing or assessment. When asked to elaborate on this definition, she gave the example below:

For example, a language exam was conducted. According to this language exam, the student's score was 70. Let's say that this student reached level B1 in the test we administered. By evaluating this student in this way and looking at his or her previous exam, how far has s/he progressed, and how far can s/he progress from now on? Is this enough for his or her field? How much work did s/he do to reach this result? What did s/he do wrong? I think it is something that we can reach with evaluations such as these (Semi-structured Interview I).

Another concept that P4-3 adequately explained was inter-rater reliability. In her explanation, she gave an example to refer to the consistency between the scores assigned by at least two raters and said, by disregarding the reliability is a matter of degree, "You and I are finalizing a test. We have concluded the test of the same person, [...] and I gave 100. You gave 50. Therefore, there is no reliability between us. It reaches such a result".

Three of the participants were able to provide appropriate answers to the concept of validity. To illustrate, P4-2 gave an example implying the importance of content validity and

emphasized, "If we ask geography in history class, I think it has no validity; I think it is low". P4-3 also expressed that she thought of validity as "the exam actually measuring the student's success" and added that if a test taker gets 100 from an exam without knowing English at all, it means that there are very serious problems with validity. When asked what could be the reasons for that, she gave the following explanation:

Because there was probably a problem with the questions asked. There may have been a problem with the measurer. The student may have cheated or done something else. The supervisor may have intervened again. So it is not entirely the student's own success. There is a problem with the reliability of the measuring instrument. There may be a problem with the impartiality of the people making the measurements. That is why, it is not valid (Semi-structured Interview I).

According to P4-5, validity means asking the students about what is taught to them. Therefore, she emphasized that an assessment should involve questions about the content taught to the students.

As for practicality, four of the senior PSETs were knowledgeable about it. They referred to some features such as being economical, time-saving, easiness in administration and scoring that are necessary for assessment tools or techniques to be practical. For instance, P4-1 pointed out that he remembered this concept from the TEE course and defined it as an assessment being prepared in line with the resources you have at that moment. Moreover, he exemplified it with an example given by the professor of the TEE course:

Our teacher gave an example like this: For example, you will prepare an exam to distinguish colors, but if there is no color printout in your region, unfortunately, we cannot use it. So usefulness. It needs to be practical. It should not pose difficulties for teachers and students, and at the same time, the factors that require it should be available around us. If it can be used, if it can be applied to students, and if we have the basic things we need to apply it, this can be shown as its usefulness (Semi-structured Interview I).

The washback effect, which means the effects of the examination on teaching, learning, and assessment, was successfully defined by P4-3 and P4-4. While P4-3 focused on classroom-based or low-stakes assessment in her definition, P4-4 considered the concept in terms of the effects of high-stakes exams. In this vein, P4-3 said that students consider the content of the teacher's assessments and direct their learning in line with these assessments with the words, "So there is a test at the end. I study accordingly. If the teacher doesn't ask me this, I won't learn it. If he asks this, I will learn this." In her explanation, P4-4 mentioned the washback effects of high-stakes exams like the Foreign Language Proficiency Test (FLPT) and said: "Studying for exams such as the FLPT can have a washback effect on our education because it is an exam-oriented study and we are actually deprived of other skills." When asked where she learned this, she said that the washback effects of the high-stakes exams were the topic of her sister's master's thesis.

Weak Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)

Some of the senior PSETs had either no knowledge or incorrect knowledge about some basic concepts, even though they had taken the MEE course in the third year and the TEELT course in the fourth year. Test is one of these terms, and two of the participants (P4-3, P4-5) define it as multiple-choice questions. Additionally, they could not provide accurate definitions related to assessment. For example, they stated that it is carried out at the end of the semester or after the testing procedure, includes reliability and validity, and involves decision-making.

Measurement, which involves the quantification of a performance with numbers or symbols according to a specific set of criteria (Bachman, 1990, p. 18), was also incorrectly defined by P4-2. According to her, it refers to the comprehensive measurement of knowledge or performance and does not include scores. Moreover, evaluation, which means decision-making based on the information collected through testing, assessment, or measurement (Bachman, 1990), could not be defined by the teacher candidates in Case 3. While P4-1 defined it as a comprehensive concept involving the scoring and analysis of the students' performance or knowledge, P4-2 commented that it includes the comparison of the student's performance with specific criteria.

Two participants gave incorrect definitions of reliability, which is the consistency of test scores obtained after two administrations. Of these, P4-3 stated that reliability means the teacher was able to assign scores accurately, and the test distinguishes the students who know from the ones who do not. Although she could not give an exact definition, P4-5 commented that reliability will be ensured if exam questions do not cause debate among students about the correct answer, as in the excerpt below:

For example, a student said that water boils at 100 degrees. The other one didn't say. This is wrong, and one should call it right, the other wrong. For example, if s/he accepts both of them as correct, it doesn't have any reliability. It seems like it's not a reliable test, then. Or, for example, if something in the test contradicts something else and they are both accepted as true, then this again has no reliability, and students will infer something anyway. It will cause a scene. Here, the wrong question was asked. This year, for example, there are news headlines because ÖSYM asked the wrong question. This undermines their credibility (Semi-structured Interview III).

Intra- and inter-rater reliability and the methods of reliability estimation were challenging concepts for the teacher candidates. Regarding these, P4-2 and P4-5 could not give any answer.

Validity is the concept that four of the senior PSETs could not accurately define. Regarding that, each of the teacher candidates who responded gave a different definition. However, the answers given by P4-2 were noteworthy because they indicated that she had some information, at least partially, about the situations that had the risk of affecting validity. To clarify, she stated that excessively challenging distractors would reduce validity.

Strong Literacy of the Key Concepts of Assessment (Third Semi-structured Interview)

According to the findings, the senior teacher trainees could also define some key concepts correctly. Therefore, their answers were put under the category of strong literacy. To clarify, three PSETs could correctly define the concept of “test.” Of these, P4-1 and P4-2 stated that it refers to a written form involving such kinds of items as multiple-choice questions, fill-in-the-blanks, and open-ended questions. Unlike these participants, P4-4 emphasized that the test may be written or include other assessment tasks.

The answers given by P4-1, P4-3, P4-4, and P4-5 indicated that they had knowledge about what measurement is. What these participants said in common is that measurement involves quantifying student performance or behavior. In addition, P4-3 said that the measurement results can also be expressed with symbols.

Reliability is also another term about which three participants have relevant and correct knowledge. All three participants emphasized that information about reliability can be obtained by administering a test to the same group twice at different times. Furthermore, P4-1 stated that the correlation between the scores obtained from the two test applications was the reliability index and that this correlation index could be between 0 and 1.

Intra- and inter-rater reliability were correctly defined by three participants (P4-1, P4-3, and P4-4). For the intra-rater reliability, they mentioned that a rater assigns scores to a language product within a while. As for the inter-rater reliability, the participants defined it by referring to the correlation between the scores awarded by at least two raters. For example, P4-3 stated this:

I think it is like IELTS in certain exams. There they have more than one person read your writing. For example, like that. If I give someone 50 and someone else gives 100, for example, then it is not reliable. If I gave 51, I would like someone else to give 53 at most (Semi-structured Interview III).

P4-3 also mentioned two methods to estimate reliability. Although she did not have detailed information about the formulas for these methods, she stated that they made calculations with the help of Excel.

To define validity, P4-1 and P4-5 emphasized its different aspects. While P4-1 emphasized content validity and predictive validity, P4-5 indirectly referred to content validity and defined validity as the evaluation of what is taught to students. In this regard, P4-5 used the following statements for validity: "I mean asking what you have taught, precisely and clearly."

The two respondents who gave correct answers about the types of validity were again P4-1 and P4-5. They both referred to content validity and predictive validity. However, P4-5 also mentioned face validity and construct validity. Regarding these types, P4-1 was able to explain what content validity is, while P4-5 was able to explain predictive validity. P4-1's definition of content validity is presented below:

Content validity is as follows: If we have taught our students up to Unit 5, the limit point where we can ask questions and make evaluations is Unit 5. Not the 6th, 7th or 8th. If a question is asked about a subject that students have never seen, this is actually a violation of content validity.

In contrast to the other concepts, all participants in Case 3 were familiar with practicality. They all emphasized economy to define usefulness. In addition, other characteristics that participants emphasized when defining usefulness were that the assessment does not require too much effort, is suitable for use with existing materials, can be used more than once, and is easily accessible.

Only two participants (P4-3 and P4-4) were able to give correct answers about the washback effect. P4-3 stated that the positive and negative effects of high-stakes examinations on classroom assessments meant washback effect. On the other hand, P4-4 defined this concept as teaching with the exam content in mind.

Knowledge about the Types of Assessment

According to the findings from the first and third semi-structured interviews, the teacher candidates in Case 3 have some misconceptions about the types of assessment, but they also have some accurate knowledge or conceptions of them. The findings in this section are also presented under the categories of weak and strong literacy, as in the previous one.

Table 47

Knowledge about the Types of Assessment

Categories	Codes (Semi-structured Interview I)	Codes (Semi-structured Interview III)
WEAK LITERACY	<i>Summative Assessment</i> • A revision of the subjects the students learned (P4-1) • Focusing on the main or key information rather than the details (P4-1) • Reading a text and summarizing (P4-2) • Measuring students' level of readiness for preparing the content of the lessons (Diagnostic assessment) (P4-3) • Informal assessment (P4-3) • Combining their scores from other exams (P4-4) • Measuring knowledge about a very general topic with 5 or 10 questions (P4-5)	<i>Summative Assessment</i> • Making inferences by looking at the general results of the assessment (P4-1) • Unable to define (P4-2) • Formal assessment involving scores (P4-5) • Requiring the student to remember what s/he has learned (P4-5) • Requiring the student to get prepared for the assessment (P4-5)
	<i>Formative Assessment</i> • Fill-in-the-blanks (P4-2, P4-5) • Formal assessment (P4-3, P4-5) • Related to constructivism (P4-1) • Producing new ideas or concepts by integrating the new knowledge with the existing one (P4-1) • Open-ended questions (P4-2) • Deciding the students' learning and success according to some specific criteria (P4-3) • True-false statements (P4-5) <i>Diagnostic Assessment</i> • Unable to define (P4-2, P4-4) <i>Dynamic Assessment</i> • Unable to define (P4-2, P4-4) • Involving all types of assessment: Summative-Formative-Diagnostic-Performance-based Assessment (P4-1, P4-3) • A continuous assessment (P4-3, P4-5) • Long-term observation of the student (P4-5)	<i>Formative Assessment</i> • Not having enough knowledge (P4-2, P4-4) <i>Diagnostic Assessment</i> • Unable to define (P4-2) <i>Dynamic Assessment</i> • Unable to define (P4-1, P4-2, P4-4) • Administering a test more than once (P4-3) • Assessment done at regular intervals (e.g., quizzes) (P4-5) <i>Alternative Assessment</i> • Fill-in-the-blanks (P4-1) • Matching (P4-1) <i>Criterion-referenced Assessment</i> • An assessment system based on a specific criterion (e.g., Criterion: Speaking, and conducting an assessment on speaking) (P4-1) • Testing whether the student learned the subject accepted as criterion (P4-5) <i>Norm-referenced Assessment</i> • Unable to define (P4-1, P4-5) • More flexible in terms of the criteria (e.g., a question having two answers) (P4-2) • Not involving scores but levels like low, mediocre, and high (P4-4)

	<i>Criterion-referenced Assessment</i> • Focusing on a specific skill or component in assessment (P4-4, P4-5) • Students giving expected answers (P4-2) <i>Norm-referenced Assessment</i> • Related to the social norms (P4-1, P4-4) • The case of ignoring minor errors (P4-2) • Unable to define (P4-5) <i>Alternative Assessment</i> • Possible to involve multiple-choice questions and fill-in-the-blanks (P4-1) • An alternative method to the actual assessment (Plan B) (P4-2) • Giving the student a second chance for assessment to prevent failure (P4-2) • Assessing grammar through speaking in addition to multiple-choice (P4-4) • Fill-in-the-blanks (P4-4) • Proficiency exam administered at preparatory school (Listening, Fill-in-the-blanks, Peer conversation for persuading) (P4-4) • Teacher giving extra points (P4-5)	
STRONG LITERACY	<i>Formative Assessment</i> • Assessment informing the teacher about what or which skill to focus on more in the subsequent lessons (P4-4) <i>Diagnostic Assessment</i> • Assessment detecting the students' strengths and weaknesses (P4-1, P4-3, P4-5) <i>Criterion-referenced Assessment</i> • Interpretation of the assessment results according to a set of criteria (P4-1, P4-3) <i>Norm-referenced Assessment</i> • Bell curve (e.g., 30% of the students failed) (P4-3) <i>Performance-based Assessment</i> • Writing a text (P4-2, P4-3, P4-4, P4-5) • Speaking about a topic (P4-2, P4-3, P4-4, P4-5)	<i>Summative Assessment</i> • Assessment done at the end of the semester (P4-3, P4-4) • Assessment done at the end of the unit (P4-4) <i>Formative Assessment</i> • Aim: Facilitating the student's learning (P4-1, P4-3, P4-5) • Tracking the student's progress (P4-5) <i>Diagnostic Assessment</i> • Assessment methods giving information about the student's strengths and weaknesses (P4-1, P4-3, P4-4) • Guiding the teacher on whether to teach again or not (P4-5) <i>Performance-based Assessment</i> • Writing a text (P4-2, P4-3, P4-4, P4-5) • Speaking (P4-3, P4-5)

• Paying attention to whether the students can accomplish the given task (Related to the task-based instruction) (P4-1) • In-class performance (P4-2) • Assessing how the student can use the four language skills (P4-3)	• An assessment during which the students being active and performing their language skills (e.g., role-play, discussion, writing) (P4-1) • Process-oriented (P4-2) • Presentation (P4-2)
<i>Traditional Assessment</i>	<i>Criterion-referenced Assessment</i>
• Multiple-choice questions (P4-1, P4-3, P4-4, P4-5) • Fill-in-the-blanks (P4-1, P4-3) • Written exams (P4-2, P4-5) • The conventional assessment methods (P4-1)	• Assessing the student according to a rubric (P4-2, P4-3) • Existence of a specific score that the student should reach (P4-4)
<i>Alternative Assessment</i>	<i>Norm-referenced Assessment</i>
• Contemporary assessment methods (P4-1) • Students being able to express their opinions and use critical thinking skills (P4-1) • Open-ended questions (P4-1) • In-class discussion between two groups (P4-1) • Writing opinions about a topic or a problem (P4-1) • Paraphrasing (P4-1) • Portfolio (P4-3) • Involving integration of language skills (P4-4) • Peer to peer conversation (P4-4) • Having an interview with someone (P4-4) • Peer assessment (P4-4) • Self-assessment (P4-4) • Focusing on speaking (P4-5) • Assessment involving more imagination (P4-5) • The student being more engaged (P4-5)	• Bell curve (P4-3) • Evaluating according to the mean score (P4-3)
	<i>Traditional Assessment</i>
	• Multiple-choice questions (P4-1, P4-2, P4-3, P4-4, P4-5) • Fill-in-the-blanks (P4-2, P4-5) • Open-ended questions (P4-2) • Pen-and-paper tests (P4-4)
	<i>Alternative Assessment</i>
	• Portfolio (P4-1, P4-2, P4-4, P4-5) • Project assignments (P4-2, P4-4, P4-5) • Peer assessment (P4-4, P4-5) • More innovative and involving variety (P4-1) • Describing a visual (P4-1) • In-class assessment (P4-1) • Recording a video (P4-2) • Authentic assessment process (P4-3) • Presentation (P4-3) • Self-assessment (P4-4) • Interview (P4-4) • Writing diaries (P4-5) • Assessments compatible with contemporary teaching methods (P4-5)

Weak Literacy of the Types of Assessment (First Semi-structured Interview)

According to the findings from the analysis of the data collected through the first semi-structured interview, the senior PSETs either have some misconceptions about the types of assessments or do not have any knowledge about them. For example, none of them could accurately define or explain summative assessment, which was the first type of assessment the researcher asked them to define. The definitions they gave involved such incorrect conceptualizations as repeating the subjects taught by the teacher, asking students to write a summary of a text, diagnosing students' readiness at the beginning of the term, calculating a score by combining the scores from other exams, and measuring knowledge about a very general topic through a limited number of questions.

Formative assessment was the other type of assessment about which four of the senior pre-service teachers lacked the necessary knowledge. Of these, P4-1 mentioned that it is related to constructivism and defined formative assessment as “the student moving to different levels based on what s/he knows.” Stating that the teacher adds new information to what the learners already know, he said that this is called scaffolding. Although he referred to the theory and the term relevant to the formative assessment, he had the thought that the learner develops his or her knowledge just by combining the existing knowledge with the new one, and that this process means the formative assessment, as in the excerpt below:

I can say that students' transition to different levels by combining their own knowledge—what they used to know and what they know now—and at the same time, since it is formative, they are able to produce new ideas by blending their old knowledge and new knowledge. [Well, can you explain the role of formative assessment there? I mean, what is the idea there, and how does formative assessment get involved there? I mean, how does it affect it?] Formative assessment is actually like this: if we assume that students start at a basic level and if students are able to blend their old knowledge with the learned knowledge and come up with new ideas, this can show that formative assessment is actually developing and can be applied. For this reason, I tried to relate formative assessment and constructivist approaches to each other based on the term (Semi-structured Interview I).

P4-2 associated the formative assessment with the knowledge of grammatical forms and expressed that it focuses on detecting whether the learners were able to write grammatically correct sentences in their answers to open-ended questions or fill-in-the-blanks with the appropriate grammatical structures. P4-5 also stated that the formative assessments are administered as written exams and involve multiple-choice questions and true-false statements. P4-3 and P4-5, like P4-2, also drew on the name of it, and both of them guessed

that it is a formal kind of assessment. Moreover, P4-3 commented that it is conducted to determine the students' level of success, as in the following comment:

Here, I saw 'formative'. I can think of something related to the school, like formal or something like that. It may be the type of assessment done at school. I believe it is stricter than a summative assessment. There are certain scoring criteria. According to these, it may be decided whether the student is successful or not. More precisely, [it may be decided] whether s/he learned those skills or not (Semi-structured Interview I).

Diagnostic assessment was also a problematic type of assessment for the participants. That is to say, two of them could not respond with any answer.

The senior PSETs needed to be more knowledgeable about DA. Three of them made comments, including their incorrect guesses about it, and the rest stated that it did not mean anything to them. To illustrate, P4-1 and P4-3 remarked the DA is a broad type of assessment and includes the previous ones like summative, formative, diagnostic, and performance-based assessment. Furthermore, P4-3 and P4-5 explained that the DA is a situation in which students are constantly assessed. Regarding this, P4-3 said: "The student may be taking this test again and again. Thus, there will be no result from just one exam. It may be that a continuous test is being done all the time". On the other hand, P4-5 expressed that this assessment would be based on observations made by the teacher.

Some of the participants also had some problems defining CRA and NRA. For the first one, P4-2 described the CRA as the test takers being able to give expected answers in the test. P4-4 and P4-5 considered the language skills or components as criteria and reported that the CRA refers to the assessment focusing on a specific skill or criterion, like writing. On the other hand, four senior PSETs needed to have the appropriate knowledge to define the NRA. While P4-5 pointed out that she did not have any comment on it, both P4-1 and P4-4 guessed it by considering the word 'norm' and commented that the NRA is an assessment related to the social norms of society. According to P4-2, the NRA is the assessor's ignoring of some errors when the learners take the assessment for the first time, and she exemplifies it like this: "For example, let's say in the first stages of the speaking exams, they say, you can make grammar mistakes; we can ignore them. I think this is the norm[-referenced assessment]".

Strong Literacy of the Types of Assessment (First Semi-structured Interview)

This section includes the findings that the senior PSETs could define or explain the types of assessment accurately. In light of these findings, however, it is to note that only some participants are familiar with the formative, diagnostic, criterion-referenced, and norm-referenced assessments. At the same time, all of them know what performance-based and traditional assessments are.

The formative assessment was successfully explained by P4-4. She addressed the role of this assessment in teachers adapting the course content and said, “For example, all of the students are very bad at listening, considering four skills. All of them or the majority of them have low scores, or we say that we need to study listening to a little more.”

Three of the senior PSETs successfully defined the diagnostic assessment. In the definition made by P4-1, for example, he pointed out, “It can be thought that it is an assessment style for diagnosing students' deficiencies and structuring them.” That is to say, the diagnostic assessment allows the teacher to detect the students' strengths and weaknesses before the teaching process and prepare their syllabuses to improve those weaknesses, and he expresses it with these words: “I can say that this is trying to learn the basic knowledge of the people participating in the course and assessment, and some syllabuses are prepared to teach the missing parts and to reinforce those subjects.”

As for the CRA, two teacher candidates, P4-1 and P4-3, explained that it aims to give results about the student's performance, considering some specific criteria. In this regard, P4-1 commented, “We can say that [the CRA] progresses in line with the syllabus and emphasizes the importance of lesson objectives. In line with this, he pointed out that in the CRA, it is assessed how much the students can attain these objectives. In a similar vein, P4-3 also drew attention to the comparison of the student's performance with the criteria and explained it with these words: “There is a criterion here. For example, knowing 50 words. For example, a test is carried out according to these criteria, and it is checked whether the student fulfills these criteria.”

NRA, which is usually acknowledged as the opposite of CRA, is the assessment type only P4-3 could explain accurately. While explaining it, she emphasized that each student's score is compared with the mean, so their rank is determined compared to other students.

Moreover, she mentioned the bell curve and the percentages used for this comparison and said: “Norm-referenced is an evaluation based on a system such as a bell curve, where, for example, thirty percent of the students fail the class.”

The performance-based, traditional, and alternative assessments are the types all senior PSETs could define or explain correctly. Regarding the first of these, the teacher trainees emphasized that the performance-based assessment involves students using language skills while carrying out authentic or simulated tasks. To illustrate, all participants referred to the assessment of speaking and writing skills and commented that it includes students talking about their experiences, making a presentation, acting out, or writing a text like an essay or a reflection paper. In addition, P4-1 stated that performance-based assessment reminded him of task-based teaching and said these:

Performance-based assessment. When I first saw performance-based assessment, I immediately thought of task-based instruction. Because it considers whether students have accomplished a task assigned to them rather than a test, this process can be painful or straightforward. But I think that performance-based assessment is this way because our main goal is to reach the stated and determined goal (Semi-structured Interview I).

All of the participants also provided answers compatible with the traditional assessment. For example, they mentioned written exams, multiple-choice questions, and fill-in-the-blanks. Furthermore, P4-5 emphasized the relationship between the grammar-translation method (GTM) and traditional assessment and expressed that the teachers using GTM in teaching English employ multiple-choice questions for measuring only the students’ competence instead of performance.

The alternative assessment was also correctly explained by the senior PSETs. In these explanations, they referred to the methods allowing the students to use the target language in a way compatible with real life. For example, interactions with peers, discussions, and presenting opinions were mentioned. Moreover, peer- and self-assessment, which enable the students either to give and receive feedback on their learning or to notice their errors by focusing on their products and guiding their learning strategies, were the methods P4-4 referred to.

Weak Literacy of the Types of Assessment (Third Semi-structured Interview)

The findings obtained from the third semi-structured interview reveal that some of the senior PSETs have either no knowledge or misconceptions about some types of assessment. To exemplify, P4-2 could not give any answers to define or explain summative assessment, while P4-1 and P4-5 made comments based on the name of it. Of these participants, P4-1 made the following explanation:

To summarize the event by making some inferences based on the assessment, based on a general whole. For example, we have an assessment criterion. We administered an assessment. We obtain an inference by looking at the general outlines of this assessment. This is a summative assessment (Semi-structured Interview III).

As for the formative assessment, P4-2 and P4-4 remarked that they did not have sufficient knowledge about it and therefore asked to continue with the next one. Moreover, P4-2 could not provide any answers to define diagnostic assessment. As the type of assessment that none of the participants in the study could define or explain correctly, dynamic assessment was also a problematic one for all the teacher candidates. Regarding that, three of them did not give any explanation. On the other hand, P4-3 made a guess based on the word ‘dynamic’ and stated that it was a "continuous measurement" and involved "performing the same test more than once." P4-5 pointed out that "I don't even remember hearing such a thing right now". Like P4-3, however, she also tried to predict this assessment by drawing on its name and stated that "it may be a measurement with a certain repetition." Moreover, she added that quizzes can be administered every three to four weeks as part of a dynamic assessment.

P4-1 commented that the alternative assessment encompasses fill-in-the-blank items and matching. P4-1 and P4-5 could not give correct explanations about CRA. For example, P4-1 defined the assessment of this concept as the assessment of the targeted language skill. On the other hand, P4-5 explained it as testing whether the learner has learned the content that is considered as a criterion and stated that "we have a criterion, and this is the main factor in doing this, we test to see if the student has learned this."

As for the NRA, two of the senior PSETs did not have any answers to define or explain it. As for CRA, P4-1 and P4-5 did not have any description or definition. On the other hand, P4-2 described this type of assessment as more flexible in terms of criteria and stated that the questions asked in this assessment can have multiple answers. P4-4, on the other hand,

stated that NRA includes classifications of students' levels as "low, medium, and high" rather than scores.

Strong Literacy of the Types of Assessment (Third Semi-structured Interview)

In contrast to the incorrect, irrelevant, or lack of knowledge about the types of assessment, some senior PSETs could also provide correct answers related to these types. Summative assessment was one of them, and two teacher candidates could give correct explanations. That is to say, both of them referred to the final exams administered at the end of the semester, while P4-4 also pointed out that these can be carried out at the end of the units with the aim of detecting what the students learn or not.

The three PSETs could give correct explanations regarding formative assessment by referring to the fact that it is done in order to contribute to students' language learning. Emphasizing that it is a process-oriented assessment, the PSETs drew attention to the importance of feedback for the formative assessment process. Moreover, P4-5 compared it with the summative assessment and added that "[summative assessment] is more formal and more score-based, while [formative assessment] is conducted more to look at further development."

As for the diagnostic assessment, the teacher candidates mentioned that it provides information about the students' strengths and weaknesses. Also, they added that the teacher can adapt his or her course content and teaching methodology according to the results of the diagnostic assessment.

Three of the senior PSETs gave correct explanations about CRA. P4-2 and P4-3 defined this type of assessment as the evaluation of student performance according to the criteria in a rubric. P4-4, on the other hand, explained it as a type of assessment in which there is a certain cut-point for students to be considered successful. P4-3 also gave correct explanations about the NRA. In these explanations, he explained that the NRA is an assessment of student achievement based on a comparison with the average and emphasized that there is a bell curve in this assessment.

Traditional and alternative assessments are the two assessments for which all participants gave correct explanations. Regarding the former, pre-service teachers stated that multiple-choice questions, fill-in-the-blank items, and open-ended questions administered as paper-and-pencil tests constitute traditional assessment. As for the alternative assessment, they emphasized that it includes portfolios, project assignments, peer assessments, and a variety of contemporary assessment methods such as presentations, interviews, and journal writing in which students use language skills.

Awareness of the Factors Affecting the Quality of Language Assessment

This section includes the findings on the senior PSETs' awareness of the factors that are likely to influence the quality of language assessments. These consist of assessment experiences, teacher qualifications, and the relationship between the language teaching and learning approaches and the assessment practices. The findings of the research reveal that the senior PSETs have a high awareness of these factors that can have an impact on the quality of the assessment. To illustrate, they stated that their assessment experiences made them hesitant to use other assessment methods, as P4-3 explained below:

It did, but I don't think it affected me well. Because I didn't see many different approaches. For example, the portfolio is a diary, etc. We didn't see many such things. I always feel like this. Because of this, when I do a portfolio or a speaking exam, the reliability of this will decrease; you know, students will not believe I am objective because objections can increase. Because students are used to it, psychologically, multiple choice questions, etc. Our teachers are probably used to it, too. So, my experience will limit me when we want to switch to a different method. But of course, I am more open to alternative techniques, such as the lessons we see at school (Focus Group Interview I).

Table 48

Awareness of the Factors Affecting the Quality of Language Assessment

STRONG LITERACY
<i>The Effects of Assessment Experiences on the Language Assessment Approach</i>
• Negative effects of the experiences on the approach to language assessment (P4-1, P4-3, P4-4) • Causing the PSET not to have knowledge about other assessment methods (P4-1) • Positive effects of the experiences on the approach to language assessment (P4-5)
<i>The Qualifications Necessary for Teachers to Conduct Quality Assessments</i>
• Being competent in language assessment (P4-1, P4-2, P4-3, P4-4) • Being competent in language skills (P4-1, P4-3, P4-4) • Being impartial (P4-1, P4-2, P4-3) • Being a good observer (P4-1) • Creativity (P4-1) • Having enough grammatical and lexical knowledge (P4-2)

-
- Being knowledgeable about the curriculum (the consistency between the assessment content and the curriculum) (P4-3)
 - Having knowledge about the student profile (P4-4)

Theoretical Knowledge about Language Assessment

- Very important (P4-1, P4-2, P4-3, P4-4, P4-5)
- Knowledge about key concepts (P4-1, P4-2, P4-3, P4-4, P4-5)
- Knowledge about assessment methods (P4-2, P4-3, P4-4, P4-5)
- Knowledge about how to prepare assessment tools (P4-2)

The Effects of Language Teaching Approach on Language Assessment Practices

- Direct reflection of teaching approach on assessment (P4-1, P4-2, P4-3, P4-4, P4-5)
 - Assessing the skills taught in the lesson (P4-3)
-

The senior student teachers also emphasized that a teacher should be proficient in language assessment in order to make quality assessments. To elaborate further, they reported that an English teacher can prepare clear instructions, different types of items, and ensure reliability and validity. Moreover, the teacher candidates pointed out that the teacher's proficiency in language skills will play an important role in the language assessments, as P4-1 wrote in his Opinion Journal.

One of the most important qualifications that an English teacher is supposed to have is language proficiency. Language proficiency is rather important so as to be able to evaluate your students. Teacher is supposed to have an advance level of English and s/he must be competent in all skills, and thus he can evaluate students (Opinion Journal II).

Furthermore, all the teacher candidates in this case noted that the theoretical knowledge about language assessment is very important. For example, all of them stated that the teacher should know key concepts like reliability, validity, practicality, and standard deviation. Also, four of them emphasized that the teacher should be knowledgeable about the features of different assessment methods while P4-2 remarked that it is important to know how to prepare various assessment tools.

Literacy in the Skills Component of LAL

Assessing the Language Skills

This section includes the findings regarding the methods, techniques, or tools the senior teacher candidates will exploit for assessing the language skills of their prospective YL and adolescent students. Emerging as a result of the analysis of data from different sources such as Opinion Journal, semi-structured interview, and scenarios, the findings are presented in terms of each language skill and group of learners in detail.

Assessing the Listening Skills

Assessing the Listening Skills of Young Language Learners

The senior PSETs will deploy both traditional, mechanical, and non-communicative items and alternative, real-like, and communicative assessment methods for assessing their future YLs' listening skills. As in other sections, the first group of practices were put under the category of weak literacy while the second one constituted the category of strong literacy.

Table 49

Assessing the Listening Skills of Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Multiple-choice questions (P4-1, P4-2, P4-4, P4-5) • Fill-in-the-blanks (P4-2, P4-4)	• Open-ended questions (P4-1, P4-3, P4-5) • Listen and act according to the instructions given by the teacher (P4-3, P4-4, P4-5) • Putting a tick, heart, or cross on the given objects, pictures, or activities according to the audio (P4-2, P4-3, P4-4) • Listen and order the pictures, sentences, and activities (sports and routines) (P4-3, P4-5) • Question-and-answer (P4-3, P4-4) • Listen and talk or write about the main idea mentioned in the audio (P4-1, P4-2) • Finding a title for the listening text (P4-1) • Finding and correcting the misinformation after watching the video (P4-5)

Weak Literacy in Assessing the Listening Skills of Young Language Learners

Multiple-choice questions and fill-in-the-blanks will be included in the written exams that will be administered by the senior PSETs. Rather than requiring the students to use their listening skills as in real-life like producing a product by drawing on the information in the audio, these types of items will require them to make some detections and give answers by choosing an option or writing a few words. Of these items, the multiple-choice questions will be the most preferred one and ask the learners to detect the topic of the listening text or what is mentioned in it. Fill-in-the-blanks, which will usually take place with the multiple-choice questions in the same exam paper, will be asked the learners to identify specific information in the audio, such as names, hours or dates.

Strong Literacy in Assessing the Listening Skills of Young Language Learners

The senior PSETs will also use more communicative or authentic methods for assessing listening skills of the YLs. In other words, they will ask their students more constructed response questions, have them act according to the given instructions, identify or order the objects mentioned by the speakers, and express own ideas briefly about the content mentioned in the audio. Moreover, P4-1 will ask his future learners to find a title for the audio they listened to while P4-5 will ask them to detect the wrong information by comparing it with the one presented in the recording and then to correct it.

Assessing the Listening Skills of Adolescent Language Learners

Listening skills of the adolescent learners of English will also be assessed through both traditional or non-communicative tools and real-like or communicative tasks. Therefore, the items to be used for non-communicative assessments constitute the category of weak literacy while the tasks or methods enabling the communicative assessment of listening skills make up the category of strong literacy.

Table 50

Assessing the Listening Skills of Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P4-2, P4-4, P4-5) • Multiple-choice questions (P4-3, P4-5) • True-False statements (P4-4)	<i>Listening and Speaking Integration</i> • Open-ended questions (P4-1, P4-3, P4-4) • Talking about the main idea mentioned in the audio (P4-1) • Information transfer questions (P4-3) • Having a conversation with a peer to find the places on the map (P4-3) • Discussing the content of the audio (P4-4) • Role-play according to the given map (P4-5) <i>Listening and Writing Integration</i> • Listening to half of a tape and writing about the rest of it (P4-1) • Dictation (P4-3) • Writing about what s/he understands from the audio (P4-4) <i>Other</i>

-
- Handing over his or her tickets to the teacher according to the announcements (P4-3)
 - Ordering the sentences by listening to the audio (P4-5)
 - Chart filling (P4-5)
 - Correcting the mistakes in the sentences by listening to the audio (P4-5)
-

Weak Literacy in Assessing the Listening Skills of Adolescent Language Learners

Three participants will exploit one or two of these items such as fill-in-the-blanks, multiple-choice questions, and true-false statements. P4-2, P4-4, and P4-5 will ask their students the fill-in-the-blanks items while they are listening to a dialog, song, or any type of audio. Regarding this, P4-2 wrote in her journal that “fill-in-the-blank is one of the best forms of assessment in listening skills”. In addition to these, P4-3 and P4-5 will exploit the multiple-choice questions in order to determine whether the learners are able to understand what they listen to by choosing among the given options. In order to use these questions, P4-5 will play a 5-minute audio recording to the students and ask them to answer these types of questions in order to determine their proficiency in listening skills at the beginning of the semester. Moreover, P4-4 wrote in her journal that she would include the true-false statements as well as the fill-in-the-blanks items in the written exams that she would administer in the future.

Strong Literacy in Assessing the Listening Skills of Adolescent Language Learners

More communicative or authentic methods will also be deployed by the senior teacher candidates in order to assess their adolescent students’ listening skills. These methods involve the integration of listening with speaking or writing skills. The first type of integration will be preferred by four of the participants, and they will require the learners to produce a speech or interact with their peers. On the other hand, the integration of listening with writing skills will consist of the learners writing a text which is produced either in a controlled way or more freely.

Assessing the Speaking Skills

Assessing the Speaking Skills of Young Language Learners

The senior teacher candidates provided communicative and authentic tasks for assessing the speaking skills of YLs of English. They will allow the students to either produce a speech on their own or have a conversation with a peer or the teacher. The first of these include their giving a short impromptu speech and a prepared presentation about a topic, talking to answer a randomly selected question, describing a picture or a word, explaining the problems they have, introducing someone in a story or recording a video of their own speeches. As for the second type of the speaking tasks, it involves the learners acting out a role given by the teacher or of their own choice, discuss a topic with their peers, talking to their teacher or peers.

Table 51

Assessing the Speaking Skills of Young Language Learners

STRONG LITERACY
<i>Production</i> • Briefly introducing himself/herself, the materials in the classroom, or visuals (P4-1, P4-2) • Making a presentation (P4-3, P4-5) • Answering the question on the randomly chosen card (P4-2, P4-5) • Describing a picture or a word (P4-2, P4-4) • Expressing their problems and what they are expected with simple sentences (P4-1) • Introducing the family members in the story told by the teacher (P4-3) • Asking a classmate to do something (Throwing a ball/Giving papers) (P4-3) • Recording short videos (P4-4)
<i>Interaction</i> • Role-play (P4-2, P4-3, P4-4, P4-5) • Discussion (P4-3, P4-4) • Having a conversation with the peers (P4-3, P4-4) • Having a short conversation with the teacher (P4-3)

Assessing the Speaking Skills of Adolescent Language Learners

Communicative and authentic assessment methods will be used by the senior PSETs in order to assess the speaking skills of their prospective adolescent students as well as those of YLs. Production and interaction are two sub-categories of these assessment methods or tasks. The tasks for production consist of the individual impromptu or prepared talks about a randomly selected topic, memories, or a reading material. On the other hand, the interaction tasks

include the role-plays and debates among the students or conversations between peers to persuade him or her and to find a place on a map.

Table 52

Assessing the Speaking Skills of Adolescent Language Learners

STRONG LITERACY
<i>Production</i>
• Making a presentation (P4-3) • Talking about the content of the reading text (P4-4) • Discussing a story after reading it (P4-4) • Answering the question on the randomly chosen card (P4-5) • Talking about their memories (P4-5) • Describing a picture or a work of art (P4-5)
<i>Interaction</i>
• Role-play (P4-2, P4-3, P4-4, P4-5) • Debate (P4-1, P4-3, P4-4) • Talking about a randomly selected or given topic (P4-2, P4-3) • Having a conversation with a peer (P4-3, P4-4) • Having an interview (P4-3)

Assessing the Reading Skills

Assessing the Reading Skills of Young Language Learners

The teacher candidates in Case 3 will assess their primary and lower secondary school students' reading skills through both traditional or non-communicative tools and alternative or communicative assessment methods. The first group of tools requires the YLs to give limited responses rather than to comprehend the short reading text and to produce any type of product. On the other hand, the second group of assessment methods asks the learners to comprehend the text and create an answer or a product based on the information in the text.

Table 53

Assessing the Reading Skills of Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Wh- questions asking for very specific information (P4-1, P4-2) • Multiple-choice questions (P4-2, P4-5) • True-False statements (P4-2)	• Simple comprehension-checking questions (P4-1, P4-3, P4-4, P4-5) • Writing a title for the given story (P4-2) • Asking critical thinking questions (P4-3) • Skimming questions (P4-4) • Scanning questions (P4-4) • Summarizing the text (P4-4)

Weak Literacy in Assessing the Reading Skills of Young Language Learners

The closed-ended wh- questions, multiple-choice questions and true-false statements constitute this category, and they are acknowledged as the indicators of weak literacy in this research. All of these items will be administered to the participants through written items in which a short text like a paragraph, dialog, or song lyrics will be given. Regarding this, for example, P4-1 and P4-2 will ask very specific questions such as the names of the characters or places in the text. Moreover, P4-2 wrote that “I usually use multiple-choice or true-false questions to assess reading skills”. In the interview, P4-5 stated that she would ask multiple-choice questions to measure whether the students are able to guess the meanings of some specific words in the text. Moreover, she accentuated that the discrete assessment of reading skills would be more rational for this group of learners.

Strong Literacy in Assessing the Reading Skills of Young Language Learners

In the strong literacy category, there are the assessment techniques allowing the learners to indicate what they understand from the text by writing a short answer or a summary, putting the mixed sentences into correct order, or determining a title for a given text. As it is obvious from these techniques, the future learners of the senior PSETs will carry out the practices which are more related to real-life than the classical items requiring them to choose among the given options. In this regard, the simple comprehension questions asking the learners to make inferences from the text and write their answer in their own words will take place in the written assessments that four of the participants will do. For example, P4-1 will give his students a 10-sentence text and ask them to write the main idea of it. Likewise, P4-2 will also exploit a way to determine whether the learners are able to comprehend the whole text and therefore her students will write a title for the short story given in the exam. Furthermore, P4-4 will also integrate reading and writing skills and ask the question of “What would you do if you were the character in the text?” and the learners will write their opinions with short and simple sentences. As another way of integrating reading with writing skills, she will also assess her students’ reading skills through a summary written by them based on the given reading text.

Assessing the Reading Skills of Adolescent Language Learners

The findings reveal that the senior PSETs will generally utilize communicative, authentic, or open-ended assessment techniques in their assessments on their future adolescent students' reading skills. However, four of them will sometimes rely on the multiple-choice questions while one of them, P4-2, will also include the true-false statements in her written exams.

Table 54

Assessing the Reading Skills of Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Multiple-choice questions (P4-1, P4-2, P4-3, P4-5) • True-False statements (P4-2) • Fill-in-the-blanks (P4-5)	• Summarizing the reading text (P4-1, P4-2, P4-4) • Comprehension-checking questions (P4-2, P4-3, P4-4) • Ordering (P4-2, P4-3) • Filling in or creating a table based on the reading (P4-3, P4-5) • Asking about the main idea of the text (P4-1) • Opinion-gap questions (P4-3) • Reading and describing a city (P4-3) • Guessing the place of the home by reading the description written by a peer (P4-4)

Weak Literacy in Assessing the Reading Skills of Adolescent Language Learners

The multiple-choice questions and true-false statements will be included in the written exams to be administered to the adolescent learners of English. Here it should be noted that P4-3 stated she would use the multiple-choice questions for assessing the reading skills of the students who would take the high-stakes exam of FLPT while the others did not make such a distinction between the groups of learners.

By the way, if the students are language class students, I would do it for that. But I would not do something like this for students in social sciences and science departments. After all, LYS 5 students will direct their future in this way and if they are not studying in beautiful cities like Ankara or Istanbul, they do not have the opportunity to take an additional course. They cannot prepare again. So, I would want them to see the same things they will see in the exam as much as possible, because the reading exams that ÖSYM actually does don't make much sense. I mean, let's say I asked critical thinking questions or something like that. But here it comes up in ÖSYM,

you do something there, you know, what could ÖSYM have put as the answer here? (P4-3, Semi-structured Interview III).

As for the true-false statements, P4-2 emphasized in the journal, interview, and scenarios that she would employ them after giving a text compatible with the learners' level. Moreover, she pointed out that she would make some purposeful changes on the statements with the aim of preventing the students from easily finding the answer. For example, she will write the dates in a way that the students have to make some easy calculations to find the correct answer or give explanations of some words in the text.

Strong Literacy in Assessing the Reading Skills of Adolescent Language Learners

In contrast to the traditional and discrete-point items, the senior teacher trainees will draw on more communicative assessment techniques. For example, the adolescent learners will write summaries based on the given text, make inferences and give answers to the constructed response items, order the events or paragraphs given in a mixed order, or create or complete a table according to the reading text. Additionally, P4-1 will give his prospective students a text and ask them to write the gist of it. In a similar way, P4-3 will aim to integrate reading and writing skills, and therefore she will give her students an article written by a traveler in order to compare the cities in Turkey. Then, she will ask her students to fill in a table comparing the characteristics of the cities and then write down where they disagree with the author. In this comparison, the students will be expected to explain their opinions and mention their memories, if any. In addition, they will write a description of a country or a city that does not exist in the table.

Assessing the Writing Skills

Assessing the Writing Skills of Young Language Learners

The findings presented in this section are related to which assessment techniques the senior PSETs will employ for assessing the lower secondary students' writing skills since there is not an explicit focus on writing in the English language curriculum for primary school (MoNE, 2018, pp. 1-48). In this vein, they reveal that the fourth-year teacher candidates will carry out assessment practices appropriate to the communicative language assessment and

conducive to the direct observation of their writing skills. Therefore, the assessment practices that they plan to do fall into the category of strong literacy as in the Table 55.

Table 55

Assessing the Writing Skills of Young Language Learners

STRONG LITERACY
• Writing a short text/paragraph to introduce oneself (P4-1, P4-2, P4-3) • Writing sentences or dialogues based on visual prompts (P4-1, P4-2) • Writing an e-mail or a letter (P4-3, P4-4) • Writing an introduction about their family members and familiar topics without much detail (P4-1) • Writing a summary of a short book (P4-3) • Writing an end to a given short story (P4-4) • Adapting a given short story (P4-4) • Preparing an invitation card (P4-4) • Preparing a menu (P4-4) • Writing a recipe (P4-4) • Answering the questions asked about a topic (e.g., planets) (P4-5)

In order to assess the YLs' writing skills, the senior PSETs will either administer writing exams in a limited period of time or give assignments that they will do individually at home. To illustrate, the learners will write a short introduction of themselves or their family members, a letter or an e-mail, sentences describing the given pictures, and a short dialog based on the given pictures in the writing exams that will be administered by P4-1, P4-2, P4-3, and P4-4. As for the assignments, P4-3 and P4-4 will integrate reading and writing skills. That is to say, they will ask the learners to read a short book and summarize it or write a different end to a given short story. Moreover, P4-4 pointed out that she would assign her students to read a short story and ask them to adapt it by considering themselves as the writer of the story. As part of these assignments, her students will also prepare such authentic materials as invitation cards, menus, and recipes.

Assessing the Writing Skills of Adolescent Language Learners

As with the YLs of English, the participants in Case 3 will make use of the communicative and authentic methods for assessing the writing skills of their adolescent learners. However, two of them, P4-2 and P4-3, pointed out that they would use dictation as a technique for writing assessment. Regarding this, it is to note that although the dictation involves the

students listening to and writing the sentences articulated by the teacher or a speaker, it does not provide the adequate data about how a learner is able to compose a text in consideration of the purpose, audience, and genre. Instead, it measures whether the students are able to hear the spoken sentences and transfer them into writing without making spelling mistakes. Therefore, this assessment technique was categorized as an indicator of weak literacy. However, in contrast to this, the methods allowing the explicit assessment of the writing skills in a discrete way or the integration of reading or listening with writing skills will also be employed by all the members of the Case 3.

Table 56

Assessing the Writing Skills of Adolescent Language Learners

STRONG LITERACY	
<i>Discrete Assessment</i>	
•	Writing an essay (P4-1, P4-3, P4-4, P4-5)
•	Writing a paragraph (P4-2, P4-3)
•	Writing a short text/story (P4-2, P4-3)
•	Defending some opinions and making explanations about these opinions (P4-1)
•	Preparing a poster (P4-2)
•	Writing about a topic by choosing among the given ones (P4-2)
•	Composition writing assignment (Writing a composition in 40 minutes) (P4-3)
•	Portfolio with the compositions written each week (P4-3)
•	Process-oriented assessment (P4-3)
•	Describing the way home (P4-4)
•	Preparing an invitation card (P4-4)
•	Writing the directions under the map prepared by the students (P4-5)
<i>Integrated Assessment</i>	
•	Writing a summary (P4-2, P4-3, P4-4)
•	Writing about a real or imaginary problematic event (P4-3, P4-4)
•	Writing a letter (P4-3, P4-4)
•	Completing a half-given story (P4-4, P4-5)
•	Giving some information and asking the students to write (P4-1)
•	Reading and describing a city (P4-3)
•	Writing about what to do in the given emergency scenario (P4-4)
•	Rewriting a story by answering some questions (P4-4)
•	Writing dialogs according to the given scenarios (P4-4)

The discrete assessment sub-category involves the assessment of writing skills without any integration with listening, reading, or speaking skills. Writing a paragraph, an essay, or a short story about a given or self-selected topic, preparing posters or invitations cards, and writing directions are the tasks in this sub-category. As for the sub-category of integrated assessment, it includes the tasks that the students will produce written texts based on the reading or listening materials. For example, they will write a summary of a book, a story, or a listening track. Moreover, they will be asked to read about a given problematic situation and write a letter of complaint imagined to be sent to the relevant organisation. Regarding this, P4-4 wrote the following in the scenarios:

I give students a situation (You are staying in a state dormitory and you are not satisfied with the dormitory you are staying in. The rooms are small, the administration is not friendly, you have complaints such as Write a petition to the Credit and Dormitories Institution Provincial Directorate stating your complaints. Explain your problems in a clear language by giving examples. State why you want to move to another dormitory). I give them the points they should pay attention to and the conjunctions etc. they should use in their writings in advance in writing (Scenarios for Foreign Language Assessment).

In the writing assessments, the adolescent learners will complete a half-given story, reflect on the content of the reading text, describe a city after reading about it, give suggestions for an emergency scenario, adapt an existing story, or create dialogs appropriate to the given scenarios. That is to say, the senior PSETs will conduct authentic and communicative assessment tasks requiring the learners to perform their writing skills as they are used in real-life.

Assessing the Language Components: Young Language Learners

Implicit assessment of the language skills are suggested in the curriculum of the primary and lower secondary schools (MoNE, 2018, pp. 1-97). Therefore, the techniques appropriate to the implicit assessment of grammar through language skills are categorized as the indicators of strong literacy while the techniques or items being conducive to explicit or discrete-point assessment are classified as the weak literacy indicators. In this context, the items that P4-4 stated that she would use for testing grammar constitute the category of weak literacy. These are classical testing items which do not require the learners to use the target grammatical structures.

As for the implicit assessment techniques, they include the use or understanding of the structures through the language skills. To illustrate, the learners will perform their speaking skills by talking about their daily routines, describing a given picture or acting out a tale. Writing will also provide the context for the implicit grammar assessments. In other words, the teacher candidates will focus on how the learners are able to use the grammar in the texts that they will write about their daily routines, likes and dislikes, or end of a tale.

In addition to these productive skills, the receptive ones will also provide the data for the implicit grammar assessment. However, the teacher trainees (P4-1 and P4-3) will assess whether the YLs are able to make the correct inferences by drawing on the form, meaning, and function of the target structures.

Table 57

Assessing Grammar: Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Fill-in-the-blanks (P4-4) • Making sentences by ordering the given words (P4-4) • True-False statements via some online applications (P4-4)	• Assessing grammar in reading (P4-1) • Assessing grammar in listening (P4-1, P4-3) ○ Understanding the instruction given by the teacher (P4-3) • Assessing grammar in writing (P4-2, P4-3, P4-4, P4-5) ○ Writing about daily routines (P4-3, P4-4, P4-5) ○ Writing an end to a tale told by the teacher and read by the students (P4-4) ○ Writing about what s/he dis/likes (P4-4) • Assessing grammar in speaking (P4-1, P4-2, P4-3, P4-4, P4-5) ○ Speaking about daily routines (P4-3, P4-5) ○ Describing what s/he sees in the picture by answering the teacher's questions (P4-3) ○ Acting out a tale (P4-4) ○ Giving short answers to the questions asked about the pictures (P4-4)

The senior teacher candidates will conduct both explicit and implicit assessments on vocabulary knowledge and use of the YLs. In order to carry out explicit assessments, for example, three of them will draw on either such items as the multiple-choice questions, fill-in-the-blanks, odd-one-out or such techniques as filling in a given table with noun, adjective, and adverb forms of the target words or writing on the exam paper the words depicted by the teacher's gestures or the pictures provided. Contrary to this, they will assess through language skills whether their prospective students are able to comprehend and use the words in written and spoken contexts. For example, P4-1 will ask some simple open-ended questions to his students while P4-2 will want them to give a brief speech about their family members. As another way of implicit assessment, P4-4 will use taboo game to assess her

students' vocabulary knowledge. Through this game they will try to explain the target words by using some relevant words and their speaking skills. As different from these, P4-5 will interact with the YLs by asking some questions that will require the learners to recognize and use the target words.

Reading and writing skills will also be utilized to conduct an assessment on vocabulary knowledge and use of YLs. In this vein, P4-1 and P4-3 will assess their students' vocabulary recognition and comprehension by reading. For example, P4-1 will give a cloze test to his students and aim at assessing whether the students can comprehend the text by recognizing the key words in text and finding the appropriate words to the written discourse. As for the assessment of vocabulary through writing, P4-2 and P4-3 will assess to what extent their students can use the words that are expected them to know and use in the written context.

Listening will also provide a context for vocabulary assessment. Regarding this, P4-3 will provide some instructions to her students and expect them to understand the instructions and act according to them. P4-4 will carry out a competition in her classroom to assess vocabulary knowledge of YLs. In this competition, the learners will listen to the audio and choose the words mentioned in it.

Table 58

Assessing Vocabulary: Young Language Learners

WEAK LITERACY	STRONG LITERACY
• Multiple-choice questions (P4-2, P4-4) • Filling in the tables with the different forms of the words (P4-1) • Fill-in-the-blanks (P4-4) • Odd-one-out (P4-1) • Writing the word on the exam paper after watching the teacher's act (P4-1)	• Assessing vocabulary in speaking (P4-1, P4-2, P4-3, P4-4, P4-5) ○ Open-ended questions (P4-1) ○ A short talk about family members (P4-2) ○ Taboo game (P4-4) ○ Question and answer (P4-5) • Assessing vocabulary in writing (P4-2, P4-3) • Assessing vocabulary in reading (P4-1, P4-3) ○ Cloze test (P4-1) • Assessing vocabulary in listening (P4-3, P4-4) ○ Understanding the TPR instructions given by the teacher (P4-3) ○ Game-like competition (YLs choosing the word on the board after listening to the audio) (P4-4)

Assessing the Language Components: Adolescent Language Learners

The adolescent or upper-secondary school learners of English will mostly take the implicit grammar assessments. However, two senior PSETs (P4-3 and P4-4) will also benefit from the traditional items to measure their knowledge of grammar. Regarding these, P4-3 emphasized that she would include the multiple-choice questions in the written exams which she would administer to the students who would take the Foreign Language Test, a high-stakes university entrance examination. On the other hand, P4-4 did not make such a distinction between the groups of learners and reported that she would exploit all of these classical items in her written exams on grammar.

Table 59

Assessing Grammar: Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Multiple-choice questions (P4-3, P4-4) • Fill-in-the-blanks (P4-4) • Sentence completion questions (P4-4)	<i>Assessing Grammar in Speaking</i> • Talking about a reading or listening text (P4-2, P4-3) • Discussion (P4-1) • Role-play (P4-3) • Question and answer (P4-5) <i>Assessing Grammar in Writing</i> • Answering open-ended questions about reading text (P4-1, P4-4) • Writing a review about a text (P4-1) • Writing a paragraph about a given topic (e.g., festivals) (P4-2) • Writing an essay (P4-3) • Writing a complaint letter (P4-3) • Writing a summary (P4-3) • Completing a partially provided story (P4-5)

The participants had the belief that the language components should be assessed implicitly through the language skills. For instance, P4-1 emphasized that grammar should be taught inductively and reminded that he read a study on the effectiveness of inductive grammar teaching. As compatible with this belief, he also accentuated that grammar should also be assessed in an implicit way as explained below:

I definitely think that grammar is actually an important structure of language, and grammar is an element of language that manifests itself in both written and spoken language. For this reason, of course, its measurement is very important and unless it is very detailed at all levels, I am

talking about the beginner level, grammar competence can be questioned. I even read a study, if I remember correctly: in two different groups, one group of students studied grammar for only one month. In another group, students analyse texts through both reading and listening. At the end of one month, when a test is done, the number of successes of the group that learnt grammar integrated rather than isolated is almost double the number of successes of the other group. In other words, we can teach grammar by incorporating it into other skills rather than teaching it in a specific way. For example, when we directly say that Simple Past is used in this way, I do not think that this is a measure of proficiency. It can be in a text. It can be in a listening text. I think that it should be evaluated in this way. In other words, I think that my evaluation criterion is that a language element should be evaluated under the roof of a basic language skill (Semi-structured Interview III).

Assessing grammar in spoken products will be preferred by four of the participants while the grammar assessment through the written products will be conducted by all of them. The future adolescent learners will either interact with each other or express their opinions based on a reading text. Discussion, role-play, and question and answer will be the methods which will be used for enabling the learners to speak in English. As for the spoken production, the learners will read or listen to a text and then give verbal comments about the content of the text.

Table 60

Assessing Vocabulary: Adolescent Language Learners

WEAK LITERACY	STRONG LITERACY
• Multiple-choice questions (P4-2, P4-3) • Fill-in-the-blanks (P4-2, P4-4) • Filling in the tables (P4-1) • Finding the wrong collocation in a given text and correcting it (P4-1) • Odd-one-out (P4-1)	<i>Assessing Vocabulary in Listening</i> • Asking for the meaning of the word in the listening text (P4-2) <i>Assessing Vocabulary in Speaking</i> • Making a short speech (P4-1, P4-2) • Role-play (P4-3) • Discussion (P4-4) • Question and Answer (P4-5) <i>Assessing Vocabulary in Reading</i> • Cloze test (P4-1, P4-4) • Asking for the meaning of the word in the reading text (P4-2) • Telling the events given in the mixed order (P4-3) <i>Assessing Vocabulary in Writing</i> • Answering open-ended questions (P4-1, P4-4) • Writing a complaint letter (P4-3, P4-4) • Writing a short text (P4-2) • Writing an e-mail (P4-5)

As in the grammar assessment, the senior teacher candidates will also rely on not only the traditional tools for explicit or discrete-point vocabulary assessment but also the communicative or authentic methods for implicit assessment of lexical items. However, it is

to note that the participants have the belief that vocabulary should be assessed implicitly as explained in the following excerpt:

Words do not exist alone in the language. The meaning of each word can change depending on the context at work. After all, I want students to use it in context. When the grammar is explicitly, it becomes memorized. I would like to do it implicitly too. So, these would be implicitly. Otherwise, I would use explicit [assessment] because I was preparing for the exam (P4-3, Semi-structured Interview III).

The multiple-choice questions, fill-in-the-blanks, filling in the tables with different forms of the words, finding the wrong collocation and correcting it, and choosing the odd word among the given ones are the items constituting the category of weak literacy.

As for the implicit assessment, the senior PSETs will mostly focus on how the adolescents are able to use vocabulary in written and verbal products and comprehend the words in the reading texts. In addition, the implicit vocabulary assessment will be conducted through the listening texts. Making a short presentation, role-play, discussion, and question and answer will be the tasks which will be deployed for assessing both speaking performance and vocabulary use. Answers to open-ended questions, complaint letters, short texts, or e-mails will provide the context for the implicit vocabulary assessments. Furthermore, cloze-tests, guessing the meanings of the words in the reading text, and telling the events given in the scrambled order in the order of their occurrence by adhering to the reading text will be the reading tasks that they will employ for assessing the learners' vocabulary comprehension.

Scoring, Interpreting, and Using the Assessment Results

Scoring the assessment results after the answers or performances elicited through various assessment techniques or tools is another important procedure in which the teacher candidates should be literate. Regarding this, the findings indicate that the senior PSETs in Case 3 will assign scores either intuitively and according to unspecific criteria or rubric or an answer key. In this regard, the first group of practices which does not involve scoring according to a pre-determined set of specific criteria constitute the category of weak literacy. In contrast to this, the scoring practices in consideration of the criteria in a rubric or the answers in an answer key are the ones falling into the category of strong literacy.

Table 61

Scoring the Answers or Performances

WEAK LITERACY	STRONG LITERACY
<i>Scoring according to Vague Criteria</i>	<i>Scoring according to a Rubric or Checklist</i>
• Giving scores out of 100 (P4-2) • Scoring according to sentence logic if there are many questions (Listening) (P4-3) • Not deducting scores for minor errors (P4-4) • Paying attention to the student's efforts (P4-4) • Having much difficulty in scoring speaking (because grammar is incorrect but the speech is intelligible) (P4-4) • Scoring according to spelling and punctuation (P4-4)	• Scoring according to a rubric or checklist (P4-1, P4-3, P4-4, P4-5) • Inter-rater reliability (P4-1) • Scoring accuracy and fluency according to a likert scale (0-5) (P4-5)
	<i>Scoring according to an answer key</i> (P4-2, P4-3, P4-4)

In the category of weak literacy, there are the practices which P4-2, P4-3, and P4-4 will carry out while scoring their prospective students' answers, products, or performances according to some criteria which are vague for the target language skills. To illustrate, P4-2 reported that she would score the written products out of 100 points and that she would pay attention to grammar and coherence in scoring. However, she did not mention she would follow the criteria in any rubric or a checklist. P4-4 will also take vague criteria into consideration while scoring. To clarify, she noted that she would pay attention to the students' efforts and ignore their grammar errors if they did not hinder intelligibility. However, she also expressed that she was not sure whether this practice was appropriate or not.

As for the strong literacy indicators, they include the practices of assigning scores to written or oral performances according to a rubric or checklist and to the answers given to the limited response items according to an answer key. In this vein, four of the participants pointed out that they would use a rubric or checklist in order to rate their students' performances in speaking and writing skills or using the language components. In these rubrics, they will include such criteria as accuracy, pronunciation, use of collocations and conjunctions, and variety of words. In addition, it should be noted that P4-1 would ask another teacher to score his future students' performances in order to ensure inter-rater reliability.

Table 62

Interpreting and Using the Assessment Results

STRONG LITERACY
<i>Interpretation and Decision about Teaching</i>
• Making changes in the teaching methods (P4-2, P4-3, P4-4) • Identifying the errors and focusing on them in the later lessons (P4-3, P4-4) • Deciding whether revision is necessary or not (P4-1) • Making an interpretation about the effectiveness of teaching (P4-2) • Deciding how often listening should be done in lessons (P4-2) • Adapting the teaching style and pace according to the results (P4-3) • Deciding whether to continue with a new topic or revise (P4-5)
<i>Discovering the Strengths</i>
• Making an interpretation about what skills the students are good at (P4-5)
<i>Discovering the Weaknesses</i>
• Detecting the common errors (P4-2, P4-3, P4-4) • Making an interpretation about why some students have low scores (P4-2) • Making interpretations about the students' weaknesses (P4-3) • Detecting the weaknesses and trying to improve them (P4-4) • Detecting the weaknesses and monitoring them with a small talk (P4-4)
<i>Giving Feedback</i>
• Giving written and oral feedback to the students (P4-3, P4-4, P4-5) • Giving the parents feedback via e-mail or by sending notes (P4-3)

As for how the senior teacher trainees will interpret and use the assessment results, the findings reveal that they will engage in the practices of interpretations and uses of assessment results which are the indicators of high literacy. These include interpretations and uses related to the effectiveness and progress of teaching and the success of the students' learning. To explain, the participants will detect their students' strengths and weaknesses and make interpretations about their reasons by drawing on the assessment results. Regarding these, they will also give feedback about them to the students and their parents through various methods. Furthermore, they will make comments on whether the teaching methods are effective or not to improve their students' language skills and thus they will make the necessary decisions to change the teaching methods in order to improve the learners' weaknesses.

Literacy in the Principles Component of LAL

In this research, the findings related to the principles component of LAL include the students' awareness of and being able to ensure ethical assessment process. Of these, the ones regarding the awareness dimension of the principles reveal that the participants in Case 3 have adequate awareness about the features of ethical assessment and the possible effects

of the stakeholders on its fairness. To be more specific, the four participants mentioned the unethical assessment experiences they had while P4-1 stated that he had not experienced any unethical assessment. As the reason for this, he referred to the use of limited response items and the compatibility between the content taught and assessed as follows:

As I mentioned in our previous interviews, we have never been subjected to an open-ended exam so far. Since there are usually multiple-choice questions, the application of ethical principles is much simpler in multiple-choice questions. Because there is a certain answer key and progress is made on it. In the concept we define as alignment, I have never encountered a question other than what was taught. Questions directly related to the subjects and units we learnt were asked. For this reason, I did not encounter any problems in terms of compliance with ethical principles (Semi-structured Interview II).

However, the other participants explained what kind of unethical assessment experiences they had. Among these, unfair scoring done by the teacher is the most frequent one that they mentioned. In this regard, they referred to both low and high scores as unfair by emphasizing that they did not correspond to their performance. As one of the reasons for this, P4-3 said that her teacher did not regard certain criteria in scoring and thus s/he rated the learners' products or performances more intuitively. In addition, she also pointed out that she witnessed the unfair scoring by her peers. That is to say, she remembered that her English teacher asked them to score the written exams and some of her peers deleted or changed their classmates' answers in order to give higher scores.

Table 63

Awareness of Ethical Assessment Process

STRONG LITERACY
<i>Unethical Assessment Experience</i>
• Getting unfairly higher or lower grades (P4-2, P4-3, P4-4) • Teacher's not showing the student the exam paper and not making any explanation about the score (P4-3) • Students' scoring each other's exam papers unfairly (P4-3) • Teacher's not scoring the answers according to specific criteria (P4-3) • Teacher's assigning higher scores to students completing the optional assignment (P4-3) • Teacher's providing students with the questions prior to the assessment (German language teacher and mathematics teacher) (P4-4) • Students' not experiencing the assessment of their language skills (P4-4) • Asking for exactly what was taught in the assessment (tertiary level) (P4-5)
<i>Features of the Ethical Assessment Process</i>
• Impartiality (P4-1, P4-2, P4-3, P4-4) • Assessing what was taught (P4-1, P4-3) • Reliability (P4-1, P4-3) • Fairness (P4-2) • Preparing very clear questions (P4-2)

-
- Distractors: Must be incorrect without any discussion (P4-2)
 - Preparing short and clear instructions (P4-2)
 - The teacher's being competent in language teaching and assessment (P4-3)
 - The necessity for the teacher to maintain objectivity (P4-3)
 - The necessity for the assessment to be recorded (P4-3)
 - The necessity for the rubric to be prepared in detail (P4-3)
 - The necessity for the students to have knowledge about the type of assessment (P4-3)
 - A well-prepared exam (P4-4)
 - The students' receiving scores proportional to their level of study (P4-5)

The Effects of Parents, Other English Teachers, and Administrators on Fairness

- Parents' pressure on the teacher (P4-1, P4-4, P4-5)
 - Administrators' criticisms (P4-1, P4-4, P4-5)
 - Other teachers' criticisms (P4-3, P4-5)
 - Differences among teachers in terms of teaching quality (P4-2)
 - Differences among teachers in the assessment methods they use (P4-2)
-

P4-3 also added that her teacher gave higher scores to the students who did a non-compulsory assignment although s/he did not inform them of the scoring. Although it was not directly related to English lessons, P4-4 reported that her German and math teachers shared the questions with the students before the assessment. Moreover, she stated that the math teacher also awarded scores without any actual assessment as given in the following excerpt:

In our German lesson, I don't want to badmouth the teacher, of course, but our teacher did not explain much like that, he did not emphasise it. He was telling it like this, but only the front row was listening to him, and when the back row didn't listen, he never told us to listen to us, and so he couldn't attract us to the lesson. For example, I don't remember anything about German in high school. That's why, after I became a language major, my maths was very bad. That teacher did something like this: he gave us the questions. He said, "Take this and do what you can, but as long as you don't cheat, it doesn't matter. Even if you do one question, do it". I'm very bad at maths too. I answered one question. I did it wrong. Then the teacher came round: "Now write down how many points you deserve." At least 70 and at most 100. I realised I had done one question, so I wrote 70. The teacher gave me 100. The one who wrote 100 got 80 (Semi-structured Interview II).

As different from the other participants, P4-5 mentioned an experience she had at university and expressed that some professors asked them to give exactly what they taught. Additionally, like P4-1, she drew attention to the limited response items and commented that she and her peers had not faced any unethical practice because the questions with exact answers had been used in the tests.

In line with these opinions about unethical assessment, the teacher candidates also mentioned various features that an ethical assessment should have. Impartiality, which is the most frequently cited feature, refers to the teacher's asking equal difficulty questions during the speaking assessments, not using content or topic appealing to an ethnic group or nation, and assigning scores by disregarding his or her personal relationship with the learner or the status of the learners' parents.

Moreover, assessing what was taught was referred by two of the participants (P4-1 and P4-3) as a feature of ethical assessment. P4-1 remarked it as one of the basic features of language assessment and said:

When it comes to language, one of the basic ethical principles is to ask for what is conveyed. In other words, for example, if a certain subject is explained, questions are asked on that subject and it is checked whether the students' goals and objectives related to that subject and unit are fulfilled. For this reason, the scope taught in ethical principles should match the scope of assessment. This is one of the important factors of ethical principles (Semi-structured Interview II).

P4-3 also emphasized this both in the journal and the interview. Regarding this, she wrote that “I measure the skills of the students by matching them with what I teach in the lesson” and thus “I measure what I have covered in”.

P4-1 also referred to the *consistency* of scoring by teacher and expressed that it was “extremely important”. Moreover, he made the explanation that the teacher should give equal scores to similar answers without including the student’s participation in the lesson in scoring.

The teacher candidates should also be aware of the parents’, other English teachers’, and administrators’ possible effects on the fairness of assessment. The findings indicate that they have the adequate awareness of these effects. It should noted, however, that their awareness is about the negative effects of these stakeholders on the fairness of assessment. That is to say, they are aware that some parents force the teachers to give higher scores or to increase the current ones. Additionally, three of the senior PSETs are aware that the administrators can criticise the teachers for giving low grades and for their choice of assessment methods. As for the effects of other English teachers, P4-3 and P4-5 mention that other English teachers may force them to exploit more traditional assessment methods or not to conduct assessments on some language skills such as listening and speaking.

As different from these, P4-2 took a different viewpoint by reporting that these negative effects of the parents, administrators, and other teachers by the other PSETs were not valid for the Turkmenistan context. However, she noted that the students in a classroom are divided into small groups of 7 or 8 students and that the students in these groups take the English lessons taught by different teachers and are therefore subjected to different assessments.

The senior PSETs will also engage in various practices to ensure ethics in assessment. Some of these practices have the potential to cause problems in ensuring ethics in assessment, while others are methods that will be effective in doing this. In this respect, while the

practices in the first group are in the category of weak literacy, those in the second category are considered as indicators that the pre-service teacher has strong literacy.

The practices in the category of weak literacy are the ones which will threaten the validity of the interpretations made from the assessment results, make it difficult to ensure the unbiased scoring, and harm the transparency of the assessment process. For example, the teacher trainees will score all the answers of a student and then continue with the other student's paper instead of scoring all the students' answers to a question, and they will assign scores in consideration of the students' efforts, which is a relative construct. Moreover, sometimes they will not share the assessment results with parents and administrators in a manner that compromises the transparency of the assessment process. Ensuring ethical online assessments is also another important concern, and some of the participants will carry out some solutions like asking different questions to the students, not conducting assessments, and administering the assessment to the students having technical problems at another time. However, these can lead to threats to validity, such as bias and cheating. Likewise, in order to assess the language skills of the students from different nations, P4-3 will follow a path that may threaten validity and give these students more time than other students because the alphabet they use in their mother tongue is different.

Table 64

Ensuring Ethical Assessment

WEAK LITERACY	STRONG LITERACY
<i>Ethical Assessment Process: Preparation, Administration, Scoring, and Interpretation</i> • Paying attention to the practicality of the assessment (P4-3) • Not scoring each student's answer to a single question before moving on to the next (P4-3) • Assigning scores based on students' efforts rather than solely prioritizing fluency (P4-4) <i>Ensuring Fairness in terms of Some Factors</i> <i>General Knowledge</i> • Not asking about the topics related to the target culture (P4-3) • Asking about topics related to culture if the students are adolescents (guessing with imagination) (P4-3) <i>Sharing the Assessment Results with Stakeholders</i> <i>Sharing the Results with Parents</i> • Sharing their child's score only in e-school (P4-3) • Sharing with the parents if asked (P4-4) • Not being sure because of apprehension about parents' reactions (P4-5) <i>Sharing the Results with Administrators</i> • No need to share directly: results are accessible on e-school (P4-1) • Sharing the results if the administrators are interested (P4-5) <i>What to Do to Ensure Fairness in Online Education</i> • Diversifying the questions related to the same topic (P4-2) • No problem with ethical assessment (P4-3) • Not administering the assessment (P4-3) • Assessing the student at another time (P4-4) <i>Ensuring Fairness in the Classroom Consisting of Students from Different Nations</i> • Giving extra time to Arab, Greek, and Russian students in writing exams (P4-3)	<i>Solution for the Unethical Assessment Experience</i> • Empathizing with the students (P4-2) • The teacher marking the exam papers: Necessary (P4-3) • Not sharing the questions with the learners (P4-4) <i>Ethical Assessment Process: Preparation, Administration, Scoring, and Interpretation</i> <i>Preparation</i> • Not choosing topics appealing to a group of learners (P4-1, P4-3, P4-4) • Taking students' personal lives into consideration (e.g., family dynamics) (P4-3, P4-4) • Choosing the topics at a similar difficulty level (P4-1) • Ensuring equal distribution of responsibilities among students by the teacher (P4-1) • Paying attention to the abundance of resources relevant to assessment topics (P4-1) • Asking questions of equal difficulty in the A & B booklets (P4-1) • Avoiding the topics unfamiliar to the students (P4-2) • Sharing the criteria with the students (P4-2) • Paying attention to the appropriateness of the content for the students' level (P4-3) • Paying attention to the compatibility of the content with what was taught (P4-3) • Ensuring assessment tools are adapted for students with disabilities (P4-3) • Not using a political or religious text that will offend a group of students (P4-3) • Preparing a very detailed rubric (P4-3) • Ensuring the appropriateness of the topic for the students (e.g., age-appropriateness, not including violence) (P4-4) • Ensuring age-appropriate listening tracks (songs or dialogues) (listen & fill-in-the blanks) (P4-4)

-
- Paying attention to the feasibility of the assessment method/tool (P4-5)

Administration

- Invigilation for the prevention of cheating (P4-1, P4-2, P4-3, P4-4, P4-5)
- Allocating equal time to the students (P4-1)
- Giving the same instructions, questions, and papers (readability) to the students (P4-1)
- Paying attention to the seating arrangement and the sound quality (listening exam) (P4-1)
- Using A-B test booklets for ensuring test security (P4-2)
- Using open-ended questions for ensuring test security (P4-2)
- Providing a quiet environment for assessment (P4-3)
- Having the student with a hearing disability sit closest to the speaker (P4-3)
- Having the visually disabled student sit in the front (P4-3)
- Recording the students' speaking performance (P4-3)
- Speaking with the students in a very intelligible way during the assessment (P4-3)
- Giving the instruction in a way that every student can hear and understand (P4-4)
- Encouraging all the students equally (P4-5)

Scoring

- Objective scoring (P4-1, P4-2, P4-3, P4-4)
- Objectivity and no discrepancy in the rubric given to the students (P4-1)
- The need for at least two raters to watch and rate the students' performance (P4-3)
- Scoring the answers carefully when not being tired (P4-4)

Interpretation

- Showing the papers to the students after the exam (P4-2)
- Interpretation of the results and making comments about the students' needs (P4-3)
- Unethical-Comparing a student's score with the mean if the students' levels are too different (P4-3)
- Comparing the student's score with his or her previous scores (P4-4)

Ensuring Fairness in terms of Some Factors

Gender

- Avoiding topics appealing to a gender (P4-2, P4-3, P4-4, P4-5)
 - Assessing what was taught in the lesson (P4-4)
-

Ethnicity/Race

- Not using the topics that can be offensive for some students (P4-3, P4-4, P4-5)
- Asking about what was taught (P4-2)
- Scoring the answers by concealing the names (Impartiality) (P4-4)
- Using a reading text about various ethnic groups (P4-4)

General Knowledge

- Asking about a very common procedure or its steps (e.g., Asking how the students come to the school) (P4-1)
- Asking about the topics relevant to the context in which the students exist (P4-1)
- Asking about the common values shared by all the students to minimize the unfairness (P4-1)
- Covering the topics in detail with video, pictures, etc. (P4-2)
- Raters not overly focusing on topic-related details (P4-3)
- The need for alignment between what was taught and assessed (P4-3)
- Providing opportunities for the students to share their opinions based on their cultural background (P4-4)
- Paying attention to the students' general knowledge (e.g., the students' daily routines in a village or small town) (P4-4)
- Teaching the students about the unfamiliar topics through similar situations before the assessment (P4-5)

Socio-economic Level

- Finding economical solutions to make the assessment more accessible (P4-1, P4-3, P4-5)
- Not requiring students to prepare a slide if there is no computer lab (P4-3)
- Considering the possible effects of socio-economic level on assessment (e.g., choosing the topics for assessment) (P4-4)

Ensuring Democracy in Language Assessment

- Getting the students' ideas about the assessment methods through questionnaires or interviews (P4-1, P4-3, P4-4, P4-5)

Ensuring Transparency in Language Assessment

Before the Assessment

- Giving the students a rubric involving the instructions, scoring procedures, time limit, and scoring criteria before the exam (P4-1)
 - Showing the students the rubric before the assessment (P4-3)
-

-
- Informing the students about the assessment conditions (P4-3)
 - Informing the students about the type and content of the assessment (P4-3)
 - Informing the students about the time, content, duration, and form of the assessment in a written or verbal way (P4-4)

During the Assessment

- Taking notes about the student's performance (P4-4)
- Using a checklist for speaking assessments (P4-4)

After the Assessment

- Giving the papers to the students after the exam (P4-2)
- Reflecting the mean score on the board (P4-3)
- Informing the students about scoring and sharing the assessment results transparently (P4-3)
- Allowing students to see their written papers and listen to their speaking recordings (P4-3)
- Being as clear as possible and answering students' questions about scoring (P4-5)

Sharing the Assessment Results with Stakeholders

Sharing the Results with Students

- Teacher-student feedback conference (P4-1, P4-2, P4-4, P4-5)
- Bringing exam papers to the classroom (P4-2, P4-3)
- Sharing the results individually through e-school (P4-1)
- Not sharing results in the classroom because of potential negative effects on student's attitude (P4-1)
- Reviewing the questions one by one in the classroom (P4-2)
- Giving general feedback on learned concepts and errors made without disclosing the students' names (P4-3)
- Giving feedback to students on areas for improvement in addition to disclosing their scores (P4-4)
- Displaying strengths and common errors on the board (P4-5)

Sharing the Results with Parents

- Talking to parents about the student's development process (P4-1, P4-2, P4-4)
- Sharing the student's score with the mean score for comparison (P4-1)

Sharing the Results with Administrators

- Giving general information about students' performance (P4-2)
 - Sharing the results with the students' names and getting feedback (P4-3)
-

Ensuring Fairness for the Students with Low Vision

- Preparing and using assessment materials written in large fonts (P4-2, P4-3, P4-4, P4-5)
- Giving the student more time for assessment (P4-2)
- Using embossed materials if necessary (P4-3)
- Student listening to the reading text via headset (P4-3)
- Getting help from the school counselor and special education teacher (P4-3)
- A student or teacher helping the student to write (P4-5)

Ensuring Fairness in the Classroom Consisting of Students from Different Nations

- Finding a common topic for all students (P4-4, P4-5)
- Not using offensive topics for discussion (e.g., refugees) (P4-3)
- Not deducting points for accent (P4-3)
- Not using topics that can be disadvantageous for some students (P4-3)
- Conducting an individualized assessment (P4-4)

What to Do to Ensure Fairness in Online Education

- Using portfolios (P4-1, P4-5)
 - Assessing with alternative assessment methods (P4-3)
 - Asking the student to record a 5-minute video (P4-3)
 - Calling the student on the phone for an assessment of speaking (P4-3)
-

However, the PSETs will also engage in practices that will ensure that the assessment process proceeds in an ethically appropriate manner. For example, they will empathize with their learners by paying attention to the impartiality of the assessment or the scoring themselves instead of having their students do it, and they will ensure the confidentiality of the questions. In addition, they will pay attention to ensure that the subjects they will use in the assessment do not provide an advantage to some students while disadvantaging others, prepare detailed rubrics, and prevent cheating. They will use rubrics, checklists and answer keys to ensure that the scoring is impartial, and will interpret the assessment results in a way that contributes to the individual development of the students by identifying areas of improvement. In addition, they will use economical methods rather than assessments that impose financial burdens on students and will seek students' views on the assessment. They will also share assessment results with students, parents, and administrators through face-to-face meetings and provide feedback on students' language development. Furthermore, they will use solutions for students with visual impairments, such as using large font sizes for the text in the assessment tools, and they will include topics in their assessments in a way that does not disadvantage or offend students from different cultures. Finally, they will use alternative assessment methods, such as portfolios or video recording, to ensure ethics in online assessments.

Reflections on the Improvement of LAL

The senior PSETs, who took both the MEE and TEELT courses, provided some reflections on the improvement of their LAL throughout the training they received in Gazi University ELT department. According to these reflections, teacher candidates made some criticisms about the education they received in the program. For example, P4-3 and P4-4 criticized the education they received within the MEE course. However, the points that these participants criticized were different. In other words, P4-3 stated that there should be more focus on the concepts of validity and reliability in this course and stated that this was not done in the MEE course. While P4-4 emphasized that she was very pleased with having prepared various items to evaluate different skills in the TEELT course, she stated that she could not say the same words for the MEE course.

I find it sufficient. In this regard [...] the training we received from the teacher, we really learned and applied different things. That's why, it contributed a lot to me. Because we also prepared questions and alternative items, we at least got an idea of how to assess them in the future. But I think I can't say the same thing about our Turkish course in the third year, because we always tried to learn numerical data, that is, things that I would never understand and that I would probably never use in my life. I mean, I can't say the same thing for them, but [...] lesson was really useful (Focus Group Interview III).

Table 65

Reflections on the Improvement of LAL

Reflections on the LAL Training
<i>Reflections on the LAL Development in General</i>
<i>Criticism</i>
• Insufficient coverage of reliability and validity in the MEE course (P4-3) • Not deriving significant benefit from the MEE course (P4-4)
<i>Suggestion</i>
• A professor from the ELT department instructing the MEE course (P4-2) • Abolishing the MEE course (P4-3) • The TEELT course being taught over two semesters (P4-3) • Focusing more on online and alternative methods (P4-3)
<i>Reflections on the TEELT course</i>
<i>Positive Reflections</i>
• Positive effects of the TEELT course (P4-3)
<i>Criticism</i>
• Not covering the alternative assessment methods (P4-2, P4-3) • Not covering ethics, reliability, and validity enough (P4-3, P4-4) • Focusing on preparing the multiple-choice questions (P4-3) • Not being competent enough in LAL because of the online course (P4-2) • Dictating the traditional assessment methods (P4-3) • Preparing multiple-choice questions for the midterm and final assessments (P4-3) • Not recommending reading resources (P4-3) • Not focusing on online assessment methods (P4-4)

However, the TEELT course was also criticized for such reasons as not focusing on alternative assessment methods, issues such as validity, reliability, and ethics, and for preparing limited response items. P4-2 and P4-3 made suggestions regarding MEE and TEELT courses. P4-2 stated that having a teacher from the ELT department in the MEE course will contribute more to their development in language assessment. P4-3, on the other hand, recommends that the MEE course should be abolished, that the TEELT course should be given in two semesters, and that this course should focus more on online and alternative assessment.

CHAPTER V

DISCUSSION AND CONCLUSION

Discussion of the Findings and Implications

The present research aimed to describe the LAL development process of the PSETs in the ELT department of Gazi University. For this purpose, the qualitative data collected through various tools during the 2020-2021 academic year were qualitatively content analysed and important findings regarding the knowledge, principles, and skills components of LAL were revealed for each case. The findings showed that a significant majority of the pre-service teachers in the three cases had experienced assessments which were usually conducted with classical assessment tools until the tertiary level. This is consistent with the findings revealed by previous research (Arslan & Üçok-Atasoy, 2020; Başok, 2017; Büyükkarcı, 2016; Haznedar, 2012, p. 45).

According to the findings related to the knowledge component of LAL, the student teachers in Case 1, consisting of 1st and 2nd graders, generally had misconceptions, inaccurate guesses, or no idea about the basic concepts and the types of assessment. For example, they stated that the concept of test means the multiple-choice items while validity refers to the teacher assessing the subject taught. Furthermore, as expected, they do not have any knowledge about the estimation method of reliability and the types of validity. On the other hand, when it comes to types of assessment, they often make interpretations based on the names of the assessments and do not provide accurate definitions or explanations of assessment types such as NRA and dynamic assessment, where it is more difficult to interpret what they are from their names. However, it is to note that the traditional assessment and alternative assessment

are the types of assessment that these participants are most familiar with and they also have experiences regarding the methods used in these assessments.

According to the findings obtained from the participants in Case 2, some of the third graders also have misconceptions about key terms like test, assessment, and measurement. However, most of the participants also have misconceptualizations about such very important concepts as reliability, validity, and washback. As for their knowledge related to the types of assessments, some of the participants in this case are not knowledgeable about types like summative, formative, diagnostic, performance-based, and criterion-referenced assessments. In addition, all of them either do not have any knowledge or have misconceptions and inaccurate guesses about dynamic assessment, while this is also the case for five of these participants and the norm-referenced assessment. In contrast to these, some of them also have knowledge about summative, formative, diagnostic, and performance-based assessment, while all of them are literate about traditional and alternative assessment.

As for the Case 3, the findings of this research indicate that the senior teacher candidates also do not have sufficient knowledge of the key concepts at the basic level. For example, two of them defined the concept of “test” as multiple-choice questions like the participants in Case 1 and 2. Additionally, most of them have low literacy in concepts such as assessment, measurement, evaluation, validity, types of validity, and washback. In contrast, some participants have knowledge about reliability, inter- and intra-rater reliability, reliability estimation strategies and types of validity. Also, it should be noted that all of these senior PSETs know what the practicality is in terms of language assessment. When it comes to their knowledge about the types of assessment, some of these PSETs have some misconceptions or no knowledge about summative, formative, diagnostic, and norm-referenced assessment types. Furthermore, the dynamic assessment is unfamiliar to all of them like other teacher candidates.

As these findings reveal, the teacher trainees do not have enough theoretical knowledge about assessment. That is to say, they either do not have any ideas about the basic concepts or have some misconceptualizations or inaccurate guesses about them, as found by the previous research (Kaplan, 2022, p. 186; Köksal, 2004). Furthermore, it should also be reported that the courses of MEE and TEELT, which were taken by the prospective teachers, were not effective enough in learning the theoretical knowledge about assessment. This case

is consistent with the one presented by the previous research focusing on the pre-service or in-service English teachers' knowledge of language assessment (Büyükkarcı, 2016; Coombe, Vafadar, & Mohebbi, 2020; Hatipoğlu, 2015; Köksal, 2004; Mede & Atay, 2017; Sarıyıldız, 2018; Şahin, 2019; Vogt & Tsagari, 2014). As a solution for this, it can be suggested that not only the teacher trainers teaching the assessment courses but also others giving methodological courses can also integrate the assessment related issues into their courses (Öz & Atay, 2017).

As for the findings regarding the skills component of LAL, the teacher trainees in Case 1 will employ both mechanical tools and communicative methods for assessing the receptive language skills of the YLs. That is to say, they will use such items as fill-in-the-blanks, multiple-choice questions, and true-false statements in order to assess their prospective YLs of English. In addition to these, they will also ask their students to do more communicative tasks like answering the open-ended questions asked about the listening or reading text, matching the pictures with the given sentences or activities, or acting according to the instructions s/he listened to. Moreover, this is generally the case for the student teachers in Case 2 and 3, although there are some differences in terms of the items or methods to be used. Therefore, it is possible to infer that the student teachers' literacies regarding the assessment of the receptive skills of YLs includes both weak and strong literacy indicators. As related to this, the previous studies revealed divergent findings. For example, Haznedar (2012) found that the in-service teachers of YLs use the classical type of items that require the students to give limited responses. In contrast to this, Kırkgöz and Ağçam (2012) reported that the in-service teachers ask more constructed response items in their written examinations.

Furthermore, the present research found that the teacher candidates in all three cases would adopt the communicative methods or techniques to assess their future YLs' speaking skills. For example, they will ask their students to give individual talks or interact with their peers or teacher through various practices like role-plays, short presentations, or question and answer activities. In this context, the teacher trainees have a high level of literacy in assessing the speaking skills of YLs of English. These findings coincide with the ones found by Yeşilçınar and Kartal (2020). However, it should be noted here that it is not known whether the pre-service teachers will actually use the methods they state that they will use to assess

their prospective YLs. As found by Öz and Atay (2017) and Yeşilçınar and Kartal (2020), there may be differences between teachers' perceptions and practices.

As for the assessment of the writing skills, the literacy of the junior student teachers has indicators at weak and strong literacy levels, while the literacy of the third and fourth year pre-service teachers is at strong level. The main reason for this is that the pre-service teachers in the first and second grades have not yet taken the "TEYL" course, and therefore they do not have enough knowledge about the English level of the students and they will not use assessment methods appropriate to their level.

The present research also yielded findings regarding the assessment of the adolescent learners' language skills. Of these, the findings related to how the teacher trainees will assess their students' receptive language skills show that they will draw on both classical, non-communicative or mechanical items and the alternative, communicative or authentic assessment methods. The first group of these assessment tools include the multiple-choice questions, fill-in-the-blanks, and true-false statements which the participants in all three cases will ask their adolescent learners in the written exams they will administer. On the other hand, they will also adopt more communicative and real-like assessment methods or techniques enabling the integration of listening or reading skills with speaking or writing skills. In this vein, it is possible to report that their literacy in assessing the receptive skills of adolescent learners has the indicators of weak and strong literacy.

As to assessing the productive skills of adolescent learners, this study found that the teacher trainees have a strong literacy regarding the assessment of speaking and writing skills. To be more precise, in the assessment of speaking skills, they will ask their students to discuss a topic with their peers, role-play, make a presentation, or express opinions about a listening or reading text. Similarly, they will carry out communicative assessments on writing by asking the adolescents to write texts like summaries, reviews, essays, e-mails, or short stories.

When all these findings are compared with the ones presented by the previous research, it can be said that they are in line with some studies (e.g., Arslan & Üçok-Atasoy, 2020; Han & Kaya 2014; Kaya, 2020). That is to say, they found that the teachers preferred the limited response items like true-false statements and matching items to assess reading skills and the communicative assessment methods like role-play to assess speaking skills. On the other hand, there is a partial consistency between the findings presented by some research (e.g.,

Mede & Atay, 2017; Ölmezer-Öztürk, 2018) since they reveal that the English teachers participating in their studies are not competent enough in assessing the language skills.

Furthermore, it should also be reported that the teacher candidates also draw on their assessment experiences while planning to assess their future students' language skills. This supports the finding found by Cheng (1999) and Smith, Hill, Cowie, and Gilmore (2014, p. 430). In their research, they revealed that the pre-service teachers' beliefs regarding the assessment were not shaped by the trainings they received on the assessment methodology and policies. However, although the findings here do not make such a strong claim, it is evident that the PSETs plan to employ the assessment techniques in their classrooms that they experienced and found beneficial.

Furthermore, as Fişne (2022, p. 126) found, the high-stakes exams like FLPT have also a negative washback effect on the preference of the assessment tools that the teacher candidates will prefer to use. Especially the language majors who are in the 12th grade and will take the FLPT at the end of the year will be exposed to the multiple-choice questions asked to measure their reading skills and knowledge of grammar and vocabulary. Therefore, it can be argued that factors such as high-stakes examinations and language policies have an impact on teachers' LAL literacy and this supports Fulcher's (2012) inclusion of historical, social political and philosophical frameworks in his description of LAL.

Moreover, as consonant with the previous studies (Hatipoğlu, 2010; Kömür, 2018; Önalın & Karagül, 2018; Sarıyıldız, 2018, p. ii; Sevimel-Şahin, 2019), the present research revealed that the teacher candidates did not have enough opportunities to practice the assessment methods during the training they took at the ELT department. In other words, the third year pre-service teachers had no opportunity to practice in the MEE course. On the other hand, the senior teacher candidates had only homework assignments for the midterm and final exams where they had to prepare limited response and free response items. However, the PSETs were aware of these shortcomings of the courses and therefore, they suggested that the courses should focus more on the practice of language assessment methods. And this was a suggestion also voiced by the pre-service teachers who participated in other studies (e.g., Kleinsasser, 2005; Malone, 2013; Şahin, 2019, p. v).

This study also revealed findings concerning the principles component of LAL. According to these findings, the teacher trainees in all cases had a high level of literacy about the characteristics of the ethical assessment process. On the other hand, the PSETs in all cases

have high and low literacy indicators on how to ensure the ethical assessment process. However, in parallel with Lam's (2015) findings, the content of assessment courses does not focus on how ethics can be ensured in the assessment process. Moreover, in contrast to the findings found by other studies (e.g., Fulcher, 2012; Giraldo & Murcia, 2018; Kremmel, Eberharter, Holzknicht, & Konrad; Yan, Fan, & Zhang, 2017), the PSETs are interested in ethics and fairness in language assessment and want to increase their literacy on these issues. Especially considering the current conditions in which Turkey is hosting a significant number of refugee students, it is vital to pay attention to and ensure ethics in language assessments.

In the light of these findings and discussion, the present research has some implications for improving the LAL of the PSETs. First of all, it can be suggested that the language assessment should be more integrated in the content of the courses in which the teacher trainees practice teaching through micro-teachings. As a finding supporting this implication, this research found that the teacher candidates who were asked to include a self-assessment section into their lesson plans preferred especially the peer- and self-assessments. Therefore, the teacher educators should emphasize in their courses that the teaching, learning, and assessment are the elements which are in interaction with each other.

Secondly, teacher trainers should choose assessment methods with the awareness that they will set an example for pre-service teachers in their assessments. In this study, the pre-service teachers stated that they would use the alternative assessment methods used by the teacher trainers in the assessments they would apply to their students in the future. Therefore, for example, they can make more use of portfolios or projects.

Thirdly, the classes of the measurement and evaluation course in education should consist of the pre-service teachers studying in the field of foreign language teaching rather than the pre-service teachers who are studying in other fields. In other words, the prospective teachers of Arabic, French, German and English should take the course together in the same class rather than with the pre-service mathematics, science or history teachers. In this way, these pre-service teachers will be able to both exchange views on foreign language assessment with each other and engage in practices that are compatible with the assessment of language skills in a communicative way. In addition, this course should be taught by teacher trainers who have knowledge not only in measurement and evaluation in education but also in the field of language education. The reason for this is that in the current situation, this course

focuses on the preparation of multiple-choice questions and the analysis of the results obtained from these tests, which is not very approved in the field of foreign language assessment.

Fourthly, the TEELT course should be taught in two semesters. In fact, this implication is put into practice during the reporting phase of the present study. The course in the first semester should include theoretical knowledge related to the foreign language assessment while the course in the second semester should focus on the how-to aspects. That is to say, it should include the practices regarding the preparation and administration of the assessment tools and interpretation of assessment results.

Furthermore, the teacher candidates can be provided with opportunities to practice assessment techniques through the teaching practice course. Within the scope of this course, pre-service teachers can practice in public schools at primary, middle, and high school levels to prepare them for teaching, and they will be able to carry out assessment and evaluation practices with these student groups. In addition, since they will have the opportunity to collaborate with in-service English teachers teaching in the practicum schools, they will also receive feedback on how these practices can be carried out in real classrooms.

Recommendations for Further Research

This research aimed to present a description of the LAL development process of the student teachers studying in the ELT department at Gazi University. For this purpose, qualitative data were collected from 20 English teacher candidates through various data collection tools and were content analyzed. As a result, important findings that will contribute to the English language teacher training program in terms of the development of LAL have been revealed. However, like every research, this study has also some limitations. Therefore, some suggestions for further research are given below.

The sample of this study consists of pre-service English language teachers at Gazi University. Therefore, there are some limitations in generalizing the findings here to other English language teaching programs. In this context, further studies can be conducted with pre-service teachers studying in English language teaching programs in other universities in order to obtain more generalizable findings.

Due to the time limitation in the doctoral program, this research includes findings from data collected from participants at four grade levels over a period of one year. Therefore, the findings on the development of pre-service teachers' LAL are a comparison of the literacy levels of different participants. However, long-term studies with the same participants over four years of teacher education will allow researchers to make more detailed comparisons of this development.

The data collected in the present research consisted of only the qualitative data. Although the qualitative data enable the researchers to obtain deeper insights into the focus of the research, it is not much feasible to collect qualitative data from a large group of participants. Therefore, it might be suggested that the researchers may also collect quantitative data through the scales in addition to the qualitative data.

As another limitation of this research, it can be stated that the participants were selected conveniently because of the pandemic. That's why, the cases were not purposefully composed of the participants with low, mediocre, and high grade point averages. As this is a factor that may affect the findings of the study, other researchers may be suggested to select participants through purposive sampling.

Furthermore, the data were not collected regarding the practicum during which the senior PSETs practice teaching and assessment because of the Coronavirus pandemic. Although the participants reported that they had not had any assessment practice during the practicum and the lesson plans they prepared for their macro teachings in the practicum were also collected and analyzed, other researchers might observe their participants in consideration of their assessment practices.

Last but not least, the English teacher training program currently includes a two-semester TEELT course on foreign language assessment. Since the teacher trainees will have more opportunities to practice language assessment during these courses, this will have an impact on the development of their LAL. Therefore, future research may be conducted to investigate how these courses contribute to the student teachers' LAL development process.

REFERENCES

- Ableeva, R., & Lantolf, J. (2011). Mediated dialogue and the microgenesis of second language listening comprehension. *Assessment in Education: Principles, Policy & Practice*, 18(2), 133-149.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Lanham, MD: American Educational Research Association.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990). *Standards for teacher competence in educational assessment of students*. ERIC Clearinhouse.
- Andrade, H., & Brookhart, S. M. (2016). The role of classroom assessment in supporting self-regulated learning. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 293-309). Cham: Springer.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12-19.
- Arslan, R. Ş., & Üçok-Atasoy, M. (2020). An investigation into EFL teachers' assessment of young learners of English: Does practice match the policy? *International Online Journal of Education and Teaching (IOJET)*, 7(2), 468-484.
- Ary, D., Jacobs, L. C., Sorensen, C. K., & Walker, D. A. (2014). *Introduction to research in education*. Belmont, CA: Wadsworth.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University.

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University.
- Bachman, L. F., & Palmer, S. A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University.
- Bademci, V. (2019). Geçerlik: Nedir? Ne Değildir?. *Eğitim ve Toplum Araştırmaları Dergisi*, 6(2), 373-385.
- Bademci, V. (2022). Correcting fallacies about validity as the most fundamental concept in educational and psychological measurement. *International e-Journal of Educational Studies*, 6(12), 148-154.
- Bailey, A., & Heritage, M. (2018). *Self-regulation in learning and the role of language and formative assessment*. Cambridge, MA: Harvard Education.
- Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557-581.
- Başok, E. (2017). *Language teaching policies and practices in the Turkish EFL context and the effects on English teachers' motivation*. Master Thesis. The University of Texas College of Education and Human Development Bicultural-Bilingual Studies, San Antonio.
- Berry, R. (2008). *Assessment for learning*. Hong Kong: Hong Kong University.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers?. *ELT Journal*, 73(2), 113-123.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. New York: Open University.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Black, P., & Wiliam, D. (2006). The reliability of assessments. In J. Gardner (Ed.), *Assessment and learning: An introduction* (pp. 119-132). London: SAGE.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), 529-549.
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126-136). Cambridge: Cambridge University.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson Education.
- Brown, H. D. (2007). *Principles of language learning and teaching* (5th edition). New York: Pearson Education.
- Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd edition).
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw Hill.
- Brown, J. D. (2021). Problems caused by ignoring descriptive statistics in language testing. In B. Lantaigne, C. Coombe & J. D. Brown (Eds.), *Challenges in language testing around the world: Insights for language test users* (pp. 15-24). Gateway East, Singapore: Springer.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University.
- Büyükkarcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Journal of Education and Instruction*, 6(3), 333-346.
- Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.
- Chaudron, C. (2000). Contrasting approaches to classroom research: Qualitative and quantitative analysis of language use and learning. *Second Language Studies*, 19(1), 1-56.
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15(3), 253-271.

- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing* (pp. 3-18). Mahwah, New Jersey: Lawrence Erlbaum.
- Cheng, L. & Fox, J. (2017). *Assessment in the language classroom: Teachers supporting student learning*. London: Palgrave.
- Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3-17). New York: Routledge.
- Coombe, C. (2018). *An A to Z of second language assessment: How language teachers understand assessment concepts*. London: British Council.
- Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, 10(3), 1-16.
- Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks: SAGE.
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches*. California: SAGE.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Ohio: Cengage Learning.
- Crocker, R. A. (2009). An introduction to qualitative research. In J. Heigham & R. A. Crocker (Eds.), *Qualitative research in applied linguistics: A practical introduction* (pp. 3-24). Hampshire: Palgrave Macmillan.
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43-56.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- Douglas, D. (2010). *Understanding language testing*. New York: Routledge.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University.

- Duff, P. A. (2008). *Case study research in applied linguistics*. New York: Lawrence Erlbaum.
- Duff, P. A. (2020). Case study research: Making language learning complexities visible. In J. McKinley, & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 144-154). London: Routledge.
- Ebadi, S., & Latif, S. V. (2015). Dynamic assessment of efl learners' listening comprehension via computerized concept mapping. *Journal of Educational Technology*, 12(2), 29-41.
- Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning*, 32(5-6), 527-555.
- Ebadi, S., & Saeedian, A. (2016). Planning future instructional programs through computerized L2 dynamic assessment. *Teaching English with Technology*, 16(4), 12-32.
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115.
- Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K. & Kyngäs, H., (2014). Qualitative content analysis: A focus on trustworthiness. *Sage Open*, 4(1), 1–10.
- Erickson, G. (2020). Finding out what learners know-and...? Reflections on teachers' language assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 29-49). Newcastle: Cambridge Scholars.
- Everitt, B. S., & Skrondal, A. (2010). *The Cambridge dictionary of statistics*. Cambridge: Cambridge University.
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(1), 1-14.
- Fişne, F. (2022). *How washback operates in macro- and micro-level dynamics of EFL teaching in Turkey: Insights from EFL teachers' testing beliefs, literacy and practices*. Doctoral Dissertation. Gazi University Institute of Educational Sciences, Ankara.
- Fosnot, C. T. (2005). Constructivism revisited: Implications and reflections. *The Constructivist*, 16(1), 1-17.

- Fosnot, C. & Perry, R. (2005). Constructivism: A psychological theory of learning. In C. Fosnot (Ed.), *Constructivism: Theory, perspectives, and practice* (pp. 8– 38). New York: Teachers College.
- Fraenkel, J. R., & Wallen, N. E., (2009). *How to design and evaluate research in education*. New York: McGraw Hill.
- Fraenkel, J. R., Wallen, N., & Hyun, H. H. (2023). *How to design and evaluate research in education*. New York: Mc Graw Hill.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- Gazi University (2020). *Information package*. Retrieved from <https://obs.gazi.edu.tr/oibs/bologna/index.aspx?lang=en>
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile Issues in Teachers' Professional Development*, 20(1), 179-195.
- Giraldo, F., & Murcia, D. (2018). Language assessment literacy for pre-service teachers: course expectations from different stakeholders. *GiST Education and Learning Research Journal*, 16, 56-77.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519-521.
- Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education*, 23(5), 531-557.
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. London: Routledge.
- Gronlund, N. E. (1998). *Assessment of student achievement* (6th ed.). Needham Heights: Allyn and Bacon.
- Han, T., & Kaya, H. İ. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education*, 4(1), 77-93.
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413-427). Germany: De Gruyter.
- Harlen, W. (2007). *Assessment of learning*. London: SAGE.

- Hatipoğlu, Ç. (2010). Summative evolution of an undergraduate ‘English language testing and evaluation’ course by future English language teachers. *English Language Teacher Education and Development*, 13, 40-51.
- Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *ELT Research Journal*, 4(2), 111-128.
- Haznedar, B. (2012). Perspectives on pre-service education of English language teachers in Turkish primary school. In Y. Bayyurt, & Y. Bektaş-Çetinkaya (Eds.), *Research perspectives on teaching and learning English in Turkey: Policies and practices* (pp. 39-57). Berlin: Peterlang.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301-307.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers?. *Language Testing*, 30(3), 345-362.
- Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. New York: Routledge.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kaplan, Y. Ü. (2022). *Development of EFL instructors’ cognitions about language assessment literacy through reflective reading and discussion*. Doctoral Dissertation, Çukurova University Institute of Social Sciences, Adana.
- Kavaklı, N., & Arslan, S. (2019). Towards a continuum from know-how to show-how for developing EFL student-teachers’ assessment literacy. *International Online Journal of Education and Teaching*, 6(1). 1261-1270.
- Kaya, T. (2020). *Exploring language assessment literacy of EFL instructors in language preparatory programs*. (Master’s Thesis). <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Kelly, B. T. (2003). Focus group interviews. In F. K. Stage & K. Manning (Eds.), *Research in the college context: Approaches and methods* (pp. 49-62). New York: Routledge.

- Kırkgöz, Y., & Ağçam, R. (2012). Investigating the written assessment practices of Turkish teachers of English at primary education. *Journal of Language and Linguistic Studies*, 8(2), 119-136.
- Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20(3), 77-102.
- Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice*, 2(2), 145-163.
- Koretz, D. (2016). Making the term 'validity' useful. *Assessment in Education: Principles, Policy & Practice*, 23(2), 290-292.
- Kozulin, A. (2003). Psychological tools and mediated learning. In A. Kozulin, B. Gindis, V. S. Ageyev & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 15-38). Cambridge: Cambridge University.
- Kozulin, A., Gindis, B., Ageyev, V. S., & Miller, S. M. (2003). Introduction: Sociocultural theory and education: Students, teacher, and knowledge. In A. Kozulin, B. Gindis, V. S. Ageyev & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 1-14). Cambridge: Cambridge University.
- Köksal, D. (2004). Assessing teachers' testing skills in ELT and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education*, 5(1), 1-11.
- Kömür, Ş. (2018). Pre-service English teachers' assessment awareness: Level of readiness for classroom practice. *The Journal of Language Teaching and Learning*, 8(1), 109-121.
- Kremmel, B., Eberharter, K., Holzknicht, F., & Konrad, E. (2017). *Enhancing language assessment literacy through teacher involvement in high-stakes test development*. Paper presented at the language testing research colloquium, Bogotá, July 2017.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.
- Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112-122.
- Lantolf, J. P. (2000). Introducing sociocultural theory. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 1-26). Oxford: Oxford University.

- Lantolf, J. P. (2007). Sociocultural source of thinking and its relevance for second language acquisition. *Bilingualism*, 10(1), 31-33.
- Lantolf, J.P., & Thorne, S.L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University.
- Lantolf, J.P., Thorne, S.L., & Poehner, M. E. (2015). Sociocultural theory and second language development. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 207-226). New York: Routledge.
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34-50.
- Lee, J., & Butler, Y. G. (2020). Reconceptualizing language assessment literacy: Where are language learners?. *TESOL Quarterly*, 54(4), 1098-1111.
- Lee, Y. W. (2015). Diagnosing diagnostic language assessment. *Language Testing*, 32(3), 299-316.
- Li, L., Gao, F., & Guo, S. (2019). The effects of social messaging on students' learning and intrinsic motivation in peer assessment. *Journal of Computer Assisted Learning*, 36(4), 439-448.
- Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, B. Gindis, V. S. Ageyev & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context* (pp. 99-118). Cambridge: Cambridge University.
- Lindgren, B. M., Lundman, B., & Graneheim, U. H. (2020). Abstraction and interpretation during the qualitative content analysis process. *International Journal of Nursing Studies*, 108, 1-6.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. New Jersey: Routledge.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344.
- Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research*, 1(2), 1-10.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40-49.

- McNamara, T. (2000). *Language testing*. Oxford: Oxford University.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, 168(1), 43-60.
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Mertler, C. A., & Campbell, C. (2005, April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.
- Morgan, C., Dunn, L., Parry, S., & O'Reilly, M. (2004). *The student assessment handbook: New directions in traditional and online assessment*. London: RoutledgeFalmer.
- Mousavi, S. A. (2002). *An encyclopedic dictionary of language testing*. Taiwan: Tung Hua.
- Ölmezer-Öztürk, E. (2018). *Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers* (Doctoral Dissertation). <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Ölmezer-Öztürk, E., & Aydın, B. (2018). Toward measuring language teachers' assessment knowledge: Development and validation of language assessment knowledge scale (LAKS). *Language Testing in Asia*, 8(1), 20-35.
- Ölmezer-Öztürk, E., & Aydın, B. (2019). Voices of EFL teachers as assessors: Their opinions and needs regarding language assessment. *Journal of Qualitative Research in Education*, 7(1), 373-390.
- Önalın, O., & Karagöl, 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.
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311-326). Cham: Springer.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402.

- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental?. *Theory into Practice*, 48(1), 4-11.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273.
- Popham, W. J. (2017). *Classroom assessment: What teachers need to know*. Boston: Pearson.
- Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6(1), 1-9.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100, 190-208.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioural Science*, 28(1), 4-13.
- Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). *Mastering modern psychological testing: Theory and methods*. Cham: Springer.
- Reynolds, C. R., Livingston, R. B., & Willson, V. (2010). *Measurement and assessment in education*. New Jersey: Pearson.
- Richardson, V. (1997). Constructivist teaching and teacher education: Theory and practice. In V. Richardson (Ed.), *Constructivist teacher education: Building a world of new understandings* (pp. 3-14). London: The Falmer.
- Roslan, R., Nishio, Y., & Jawawi, R. (2022). Analyzing English language teacher candidates' assessment literacy: A case of Bruneian and Japanese universities. *System*, 111, 1-9.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, 11, 1-13.
- Sarıyıldız, G. (2018). *A study into language assessment literacy of preservice English as a foreign language teachers in Turkish context* (Master's Thesis). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Savignon, S. J. (1991). Communicative language teaching: State of the art. *TESOL Quarterly*, 25(2), 261-278.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago, IL: Rand McNally.

- Sevimel-Şahin, A. (2019). *Exploring foreign language assessment literacy of pre-service English language teachers* (Doctoral Dissertation). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Selvi, A. F. (2019). Qualitative content analysis. In J. McKinley & H. Rose, *The Routledge handbook of research methods in applied linguistics* (pp. 440-452). London: Routledge.
- Shrestha, P. (2012). *Dynamic assessment of academic writing for business studies* (Doctoral Dissertation). Retrieved from <https://oro.open.ac.uk/61606/1/13835912.pdf>
- Shrestha, P., & Coffin, C. (2012). Dynamic assessment, tutor mediation and academic writing development. *Assessing Writing*, 17(1), 55-70.
- Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of Research in Education*, 39(1), 215-252.
- Smith, L. F., Hill, M. F., Cowie, B., & Gilmore, A. (2014). Preparing teachers to use the enabling power of assessment. In C. Wyatt-Smith, V. Klenowski & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 418-445). Dordrecht: Springer.
- Stiggins, R. (1991). Assessment literacy. *The Phi Delta Kappan*, 72(7), 534-539.
- Stiggins, R. (1995). Assessment literacy for the 21st century. *The Phi Delta Kappan*, 77(3), 238-245.
- Stiggins, R. J. (1999). Assessment, student confidence, and school success. *The Phi Delta Kappan*, 81(3), 191-198.
- Şahin, S. (2019). *An analysis of English language testing and evaluation course in English language teacher education programs in Turkey: Developing language assessment literacy of pre-service EFL teachers* (Doctoral Dissertation). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412.
- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2018). *İngilizce Dersi Öğretim Programı (İlkokul ve Ortaokul 2, 3, 4, 5, 6, 7 ve 8. sınıflar)*. [English language teaching program (Primary and Secondary schools the 2nd, 3rd, 4th, 5th, 6th, 7th and 8th grade)]. Ankara: MEB.

- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2018). *Ortaöğretim İngilizce Dersi (9, 10, 11 ve 12. Sınıflar) Öğretim Programı*. [English Language Teaching Program for secondary schools (The 9th, 10th, 11th and 12th grade)]. Ankara: MEB.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Topping, K. (2018). *Using peer assessment to inspire reflection and learning*. New York: Routledge.
- Vadeboncoeur, J. A. (1997). Child development and the purpose of education: A historical context for constructivism in teacher education. In V. Richardson (Ed.), *Constructivist teacher education: Building a world of new understandings* (pp. 15-38). London: The Falmer.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770.
- Von Glasersfeld, E. (1989). Constructivism in education. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (pp. 114-117). Oxford: Pergamon.
- Von Glasersfeld, E. (2005). Thirty years constructivism. *Constructivist Foundations*, 1(1), 9-12.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University.
- Yan, X., Fan, J., & Zhang, C. (2017). *Understanding language assessment literacy profiles of different stakeholder groups in China: The importance of contextual and experiential factors*. Paper presented at the language testing research colloquium, Bogotá, July 2017.
- Yeşilçinar, S., & Kartal, G. (2020). EFL teachers' assessment literacy of young learners: findings from a small-scale study. *Journal of Theoretical Educational Science*, 13(3), 548-563.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.

- Wertsch, J. V. (1993). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge: Harvard University.
- William, D., & Thompson, M. (2017). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment* (pp. 53-82). New York: Routledge.
- Wylie, E. C., & Lyon, C. J. (2020). The role of technology-enhanced self-and peer assessment in formative assessment. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 170-191). New York, NY: Routledge.
- Xu, Y., & Brown, G. T. (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., & 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.
- Yastıbaş, A. E. (2018). Language assessment literacy of Turkish EFL instructors: A multiple-case study (Doctoral Dissertation). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Yılmaz-Yakışık, B. (2012). *Dynamic assessment of ELT students' speaking skills* (Doctoral Dissertation). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Yin, R. K. (2003). *Applications of case study research*. Thousand Oaks: SAGE.
- Yin, R. K. (2018). *Case study research and applications: Design and Methods*. California: SAGE.



APPENDICES

Appendix 1. Bilgilendirilmiş Gönüllü Olur Formu

BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU

Bu form, Gazi Üniversitesi İngiliz Dili Eğitimi Anabilim dalında öğrenim gören hizmet öncesi İngilizce öğretmenlerine yönelik olarak “The Description of Pre-service EFL Teachers’ Language Assessment Literacy Development Process (Hizmet Öncesi İngilizce Öğretmenlerinin Yabancı Dilde Ölçme Değerlendirme Okuryazarlığı Geliştirme Sürecinin Betimlenmesi)” isimli çalışmaya katılacak kişilerin gönüllü onayını almak amacıyla hazırlanmıştır.

Doktora tezim kapsamında yürüttüğüm bu çalışma, İngilizce öğretmenliği programında öğrenim gören hizmet öncesi İngilizce öğretmenlerinin yabancı dilde ölçme ve değerlendirme okuryazarlığına ilişkin gelişim sürecini araştırmayı amaçlamaktadır. Bahsi geçen çalışmayı yürütmek amacıyla, İngilizce öğretmeni adaylarıyla yarı yapılandırılmış görüşmeler gerçekleştirecek, öğretmen adaylarına yabancı dilde ölçme ve değerlendirme okuryazarlığı gelişim sürecine ilişkin günlükler yazdıracak, çalışmaya katılacak öğretmen adayları tarafından gerçekleştirilen mikro öğretimleri gözleyecek ve bu katılımcılar tarafından hazırlanan ders planlarını ve ilgili ödevleri inceleyeceğim.

Yukarıda kapsamı belirtilen bu çalışmaya katılmak tamamen gönüllülük esasına dayanmaktadır. Bu nedenle, çalışmaya katılan öğretmen adayları istedikleri zaman mazeret belirtmeksizin çalışmadan ayrılma hakkına sahiptir. Çalışma kapsamında araştırmacı tarafından elde edilecek veriler yalnızca bilimsel amaçlar için kullanılacaktır ve katılımcılara ait kişisel bilgiler gizli tutulacaktır. Ayrıca, çalışmaya katılmanız için sizden herhangi bir ödeme talep edilmeyecektir. Belirtilen koşullar dışında, çalışmaya ilişkin daha fazla bilgi edinmek isterseniz araştırmacı ile iletişime geçebilirsiniz.

Çalışmanın içeriği hakkında yeterince bilgi edindiğinizi, çalışmaya tamamen gönüllü olarak katıldığınızı ve çalışmadan hiçbir mazeret belirtmeksizin ayrılabilceğinizi beyan ediyorsanız aşağıda size ayrılan kısımları doldurabilirsiniz. Bu çalışmanın yürütülmesini sağlayacak kıymetli katılımlarınız için teşekkür ederim.

Arş. Gör. Murat ŞÜKÜR

Katılımcının

Adı-Soyadı:

Sınıfı:

Cinsiyeti:

Yaşı:

İmza:

Appendix 2. Questions for the Opinion Journal I

Lütfen, aşağıda verilen sorulara tamamen kendi düşüncelerinizi ve deneyimlerinizi yansıtacak şekilde detaylı cevap veriniz. Bu günlük kapsamında vermiş olacağınız özel bilgiler tamamen gizli tutulacak olup üçüncü şahıslarla izniniz olmadan hiçbir şekilde paylaşılmayacaktır. Zaman verip, emek harcadığınız için teşekkür ederim.

1. Şimdiye kadarki eğitim hayatınız boyunca İngilizce dil becerileri bakımından hangi yöntemlerle ve nasıl değerlendirildiniz? (Örn. çoktan seçmeli sorular, boşluk doldurma soruları, açık uçlu sorular, gözlem vb.) Yaşamış olduğunuz bu değerlendirmeler hakkındaki görüşleriniz nelerdir?
2. Yaşamış olduğunuz bu değerlendirmeleri göz önünde bulundurduğunuzda, İngilizce öğretmeninizin yerinde olsaydınız neleri aynı şekilde uygulardınız ve neleri değiştirdiniz? Nedenleriyle açıklayınız.

Appendix 3. Questions for the Opinion Journal II

Lütfen, aşağıda verilen sorulara tamamen kendi düşüncelerinizi ve deneyimlerinizi yansıtacak şekilde detaylı cevaplar veriniz. Bu günlük kapsamında vermiş olacağınız özel bilgiler tamamen gizli tutulacak olup üçüncü şahıslarla izniniz olmadan hiçbir şekilde paylaşılmayacaktır. Zaman verip, emek harcadığınız için teşekkür ederim.

1. Bir İngilizce öğretmenin, dil becerilerine yönelik nitelikli değerlendirmeler yapabilmesi için hangi özelliklere/niteliklere sahip olması gerekmektedir? Nedenlerle açıklayınız.
2. Bir İngilizce öğretmenin yaptığı değerlendirmelerin niteliğini hangi faktörler etkileyebilir? Bu faktörlerin yapılan değerlendirmelerin niteliğini nasıl etkilediklerini açıklayınız.



Appendix 4. Questions for the Opinion Journal III

Lütfen, aşağıda verilen sorulara tamamen kendi düşüncelerinizi ve deneyimlerinizi yansıtacak şekilde detaylı cevaplar veriniz. Bu günlük kapsamında vermiş olacağınız özel bilgiler tamamen gizli tutulacak olup üçüncü şahıslarla izniniz olmadan hiçbir şekilde paylaşılmayacaktır. Zaman verip, emek harcadığınız için teşekkür ederim.

1. İngilizce öğretmeni olduğunuzda öğrencilerinizi dil becerileri bakımından değerlendirirken etik açısından hangi hususlara dikkat edersiniz?
2. Yapacağınız bu değerlendirmelerde öğrencileriniz arasında adaleti sağlamak için neler yaparsınız?



Appendix 5. Questions for the Opinion Journal IV

Lütfen, aşağıda verilen sorulara tamamen kendi düşüncelerinizi ve deneyimlerinizi yansıtacak şekilde detaylı cevaplar veriniz. Bu günlük kapsamında vermiş olacağınız özel bilgiler tamamen gizli tutulacak olup üçüncü şahıslarla izniniz olmadan hiçbir şekilde paylaşılmayacaktır. Zaman verip, emek harcadığınız için teşekkür ederim.

1. İngilizce öğretmeni olduğunuzda öğrencilerinizi dört dil becerisi ve dil bilgisi, kelime ve telaffuz gibi dil ile ilgili bileşenler bakımından nasıl değerlendirmeyi planlıyorsunuz? Detaylı olarak açıklayınız.
2. Yapacağınız bu değerlendirmelere teknolojiyi nasıl entegre edebilirsiniz? Örnekler vererek açıklayınız.



Appendix 6. Questions Asked in the First Semi-structured Interview

1. Ölçme-değerlendirme alanı açısından aşağıdaki kavramları nasıl tanımlarsınız?
 - a. Test
 - b. Assessment (Düzey belirleme, bellilendirme, ölçme-değerlendirme)
 - c. Measurement (Ölçme)
 - d. Evaluation (Değerlendirme)
2. Özetleyici ölçme-değerlendirme (summative assessment) kavramını nasıl tanımlarsınız?
3. Biçimlendirici ölçme-değerlendirme (formative assessment) kavramını nasıl tanımlarsınız?
4. Tanılayıcı ölçme-değerlendirme (diagnostic assessment) kavramını nasıl tanımlarsınız?
5. Performans temelli ölçme-değerlendirme (performance-based assessment) kavramını nasıl tanımlarsınız?
6. Dinamik ölçme-değerlendirme (dynamic assessment) kavramını nasıl tanımlarsınız?
7. Geleneksel ölçme-değerlendirme (traditional assessment) kavramını nasıl tanımlarsınız?
8. Geleneksel ölçme-değerlendirmede kullanılan yöntemler veya araçlar nelerdir?
9. Alternatif ölçme-değerlendirme (alternative assessment) kavramını nasıl tanımlarsınız?
10. Alternatif ölçme-değerlendirmede kullanılan yöntemler veya araçlar nelerdir?
11. Kriter temelli ölçme-değerlendirme (criterion-referenced assessment) kavramını nasıl tanımlarsınız?
12. Norm-temelli ölçme-değerlendirme (norm-referenced assessment) kavramını nasıl tanımlarsınız?
13. Güvenirlik (reliability) kavramını nasıl tanımlarsınız?
14. Intra-rater reliability kavramını nasıl tanımlarsınız?
15. Inter-rater reliability kavramını nasıl tanımlarsınız?
16. Güvenirliği sağlamak için kullanılan yöntemler nelerdir?
17. Geçerlik (validity) kavramını nasıl tanımlarsınız?
18. Geçerlik türleri nelerdir?
19. Geçerliği sağlamak için kullanılan yöntemler nelerdir?

20. Kullanışlılık (practicality) kavramını nasıl tanımlarsınız?

21. Washback effect kavramını nasıl tanımlarsınız?



Appendix 7. Questions Asked in the Second Semi-structured Interview

1. Sizce etik açısından uygun bir ölçme-değerlendirme süreci nasıldır?
2. Eğitim hayatınızda bugüne kadar yaşamış olduğunuz değerlendirmeleri düşündüğünüzde bu değerlendirmelerde etik ilkelere uygunluk bakımından herhangi bir sorunla karşılaştınız mı? Cevabının “Evet” ise, bu sorunu açıkla mısınız?
3. Öğrencilerinizin dil becerilerini değerlendirmek için kullanacağınız ölçme araçlarını hazırlarken etik açısından hangi noktalara dikkat edersiniz?
4. Hazırladığınız ölçme araçlarını uygularken etik açısından hangi noktalara dikkat edersiniz?
5. Ölçme araçları vasıtasıyla elde ettiğiniz sonuçları puanlarken ve yorumlarken etik açısından hangi noktalara dikkat edersiniz?
6. Öğrencilerinizi dil becerileri bakımından değerlendirirken aşağıdaki faktörler açısından öğrencileriniz arasında adaleti nasıl sağlayabilirsiniz?
 - a. Cinsiyet
 - b. Etnik kimlik/Irk
 - c. Genel kültür
 - d. Sosyo-ekonomik düzey
7. Dilde ölçme-değerlendirme açısından şeffaflık (transparency) nedir?
8. Şeffaflığa ilişkin ifade ettiğiniz tanımdan yola çıkarak ölçme-değerlendirme sürecinde şeffaflığı nasıl sağlarsınız?
9. Değerlendirme sonuçlarını öğrencilerinize nasıl bildirirsiniz?
10. Değerlendirme sonuçlarını velilere bildirmeyi tercih eder misiniz? Cevabınız “evet” ise nasıl bildirirsiniz?
11. Değerlendirme sonuçlarını diğer İngilizce öğretmenlerine bildirmeyi tercih eder misiniz? Cevabınız “evet” ise nasıl bildirirsiniz?
12. Değerlendirme sonuçlarını yöneticilere bildirmeyi tercih eder misiniz? Cevabınız “evet” ise nasıl bildirirsiniz?
13. Aşağıdaki ölçme-değerlendirme uygulamalarının öğrencilerinizin dil öğrenme süreci üzerinde ne tür etkilerinin olacağını düşünüyorsunuz?
 - a. Sınıf bazında yapılan ölçme-değerlendirme uygulamaları
 - b. Ulusal çapta yapılan dil sınavları (LGS/YDT/YDS/TOEFL/IELTS)

Appendix 8. Questions Asked in the Third Semi-structured Interview

1. Öğrencilerinizin dil becerilerini her beceri birbirinden bağımsız (discrete) olacak bir şekilde değerlendirmeyi mi, yoksa tümleşik (integrated) olacak bir şekilde değerlendirmeyi mi tercih edersiniz? Neden?
2. Öğrencilerinizin İngilizce dinleme becerilerini nasıl değerlendirmeyi düşünüyorsunuz? Öğretim kademelerine göre açıklayınız.
3. Öğrencilerinizin İngilizce konuşma becerilerini nasıl değerlendirmeyi düşünüyorsunuz? Öğretim kademelerine göre açıklayınız.
4. Öğrencilerinizin İngilizce okuma becerilerini nasıl değerlendirmeyi düşünüyorsunuz? Öğretim kademelerine göre açıklayınız.
5. Öğrencilerinizin İngilizce yazma becerilerini nasıl değerlendirmeyi düşünüyorsunuz? Öğretim kademelerine göre açıklayınız.
6. Dil bileşenlerini (kelime ve dil bilgisi) dil becerileri ile birlikte ve dolaylı/üstü kapalı (implicitly) bir şekilde değerlendirmeyi mi, yoksa her birini ayrı ayrı ve açık/doğrudan (explicitly) bir şekilde değerlendirmeyi mi tercih edersiniz? Neden?
7. Öğrencilerinizin İngilizce dil bilgisi yeterliğini nasıl değerlendirmeyi düşünüyorsunuz? Öğretim kademelerine göre açıklayınız.
8. Öğrencilerinizin kelime bilgisini ve kullanımını nasıl değerlendirmeyi düşünüyorsunuz? Öğretim kademelerine göre açıklayınız.
9. Öğrencilerinizin İngilizce telaffuz yeterliğini değerlendirmeyi düşünüyor musunuz? Eğer cevabınız “Evet” ise, bu yeterliği nasıl değerlendirmeyi düşünüyorsunuz? Öğretim kademelerine göre açıklayınız.
10. Çocuk yaşıta İngilizce öğrenen öğrencilerinizin İngilizce dil becerilerini nasıl değerlendirmeyi düşünüyorsunuz?
11. Öğrencilerinizin İngilizce dil becerilerine yönelik yaptığınız ölçme işlemlerinden sonra ortaya çıkan cevapları, ürünleri veya performansları nasıl puanlamayı düşünüyorsunuz?
12. Öğrencilerinizin İngilizce dil becerilerine yönelik yaptığınız ölçme işlemlerinden elde ettiğiniz sonuçları nasıl yorumlamayı düşünüyorsunuz?

Appendix 9. Questions Asked in the Focus Group Interview I

1. Bugüne kadarki ölçme-değerlendirme deneyimlerinizi incelediğimde İngilizce öğretmenlerinizin değerlendirmelerde genellikle çoktan seçmeli sorular, boşluk doldurma soruları, doğru yanlış ifadeleri ve açık uçlu sorular kullandıklarını görüyorum. Sizce öğretmenlerinizin bu değerlendirme yöntemlerini kullanma nedenleri nelerdir?
2. İngilizce dil becerileri bakımından bugüne kadar yaşamış olduğunuz değerlendirmeler dilde ölçme ve değerlendirmeye yönelik yaklaşımınızı etkiledi mi? Cevabınız “evet” ise, bu değerlendirmelerin yaklaşımınızı ne kadar ve nasıl etkilediğini düşünüyorsunuz?
3. Günlüklerinizde ve ilk bireysel görüşmedeki yanıtlarınızda “derste ne öğretildi ise onun değerlendirilmesi gerekir” düşüncesiyle sıklıkla karşılaştım. Bu bağlamda, sizce bir İngilizce öğretmenin İngilizceyi öğretmeye ve öğrenmeye yönelik yaklaşımı, ölçme ve değerlendirme uygulamalarını nasıl etkiler?
4. İkinci günlüklerinizde bir İngilizce öğretmenin ölçme ve değerlendirme hakkında teorik bilgiye sahip olması gerekir görüşü ile karşılaştım. Sizce sahip olunması gereken bu teorik bilgiler nelerdir?
5. Bir İngilizce öğretmenin uyguladığı değerlendirmelerin türü ve niteliği üzerinde görev yaptığı okul, veliler ve yöneticiler gibi unsurların sizce herhangi bir etkisi var mıdır? Cevabınız “evet” ise, bu unsurların ne kadar ve nasıl bir etkisi olabilir?

Appendix 10. Questions Asked in the Focus Group Interview II

1. Günlüklerinizde ve görüşmelerinizde ölçme-değerlendirme sürecinin adil olması gerektiği görüşü hakimdi. Bu bakımdan, öğrencilerinizin dil becerilerini değerlendirirken hangi ölçme-değerlendirme araçlarının veya uygulamalarının öğrencileriniz açısından daha adaletli sonuçlar vereceğini düşünüyorsunuz?
2. Öğrencilerinizi dil becerileri ve dil bileşenleri bakımından değerlendirirken sizce hangi dil becerisi/becerileri veya dil bileşeni/bileşenleri bakımından adil olmanın daha kolay veya zor olacağını düşünüyorsunuz? Neden?
3. Bazı günlüklerde dış etmenlerin de ölçme-değerlendirme sürecinde etik açısından etkili olabileceği görüşü ile karşılaştım. Sizce veliler, diğer İngilizce öğretmenleri ve yöneticiler gibi paydaşların bir İngilizce öğretmenin yapacağı değerlendirmelerde etik ilkelere uymak bakımından herhangi bir etkisi var mıdır? Cevabınız “evet” ise, bu paydaşların değerlendirmeyi etik açıdan ne kadar ve nasıl etkileyeceğini düşünüyorsunuz?
4. Çevrimiçi eğitim şartlarını göz önünde bulundurduğunuzda, öğrencilerinizin dil becerilerini değerlendirirken etik açıdan uygun bir ölçme-değerlendirme süreci oluşturmak için neler yaparsınız?
5. İngilizce öğretmeni olduğunuzda dilde ölçme-değerlendirme uygulamalarının daha demokratik olabilmesi için neler yapabilirsiniz?

Appendix 11. Questions Asked in the Focus Group Interview III

1. Görüşmelerimizde, günlüğünüzde ve senaryolarda ölçme-değerlendirme sonuçlarından faydalanarak öğrencilerinize dönütler vereceğinizi görüyorum. Öğrencilerinize vereceğiniz bu dönütlerin nasıl olacağı hakkında detaylı bilgi verir misiniz?
2. Yabancı dilde ölçme-değerlendirme okuryazarlığı gelişiminde öğretmen eğitimcilerin (akademisyenlerin) nasıl roller üstlenmelerini ve hangi niteliklere sahip olmaları gerektiğini düşünüyorsunuz?
3. Gazi Üniversitesi İngiliz dili eğitimi programında yabancı dilde-ölçme değerlendirme okuryazarlığı bakımından gelişiminize ilişkin alacağınız/aldığınız eğitim hakkında beklentileriniz/görüşleriniz nelerdir?
4. Sizce Yabancı Dil Öğretiminde Ölçme ve Değerlendirme/İngilizce Öğretiminde Sınav Hazırlama dersinin içeriği nelerden oluşmalıdır?

Appendix 12. Scenarios for Foreign Language Assessment (Turkish Version)

Yabancı Dilde Değerlendirme Senaryolarının (YDDS) amacı, aşağıda verilen durumlarla karşılaştığınızda öğrencilerinizi nasıl değerlendirebileceğinize ilişkin bilgi toplamaktır. Bu veri toplama aracı 13 farklı senaryo ve 33 adet soru içermektedir. Lütfen aşağıda verilen senaryoların her birini dikkate alarak sorulara mümkün olduğunca ayrıntılı cevaplar veriniz.

1. Yabancı dil olarak İngilizce öğrenenlerden oluşan yeni bir sınıfınız olduğunu varsayalım. Ancak, öğrencilerinizin İngilizce seviyelerini bilmiyorsunuz ve bu öğrencilere ders vermeye başlamadan önce onlar hakkında bilgi edinmeniz gerekiyor. İngilizcedeki güçlü ve zayıf yönlerini tespit etmek için öğrencilerinizi nasıl değerlendirirsiniz?
2. A1 İngilizce seviyesinde çocuk yaştaki öğrencilerden oluşan bir sınıfınız var. Bu öğrencilerden (a) sınıftaki nesnelerin isimlerini ve renklerini belirlemeleri ve (b) İngilizce verilen kısa ve basit sözlü yönergeleri ve talimatları izlemeleri beklenmektedir. Bu durumda, öğrencilerinizin dinleme becerilerini değerlendirmek istiyorsunuz.
 - Öğrencilerinizin dinleme becerilerini nasıl değerlendirirsiniz?
 - Dil bileşenlerini (kelime ve gramer) nasıl değerlendirirsiniz?
3. Bir ilkokulda çocuk yaştaki öğrencilere (A1 Seviyesi) İngilizce öğretiyorsunuz. Öğrencilerinizden İngilizce kısa ve basit ifadeler kullanarak ailelerindeki kişileri tanıtmaları ve bu kişilerin fiziksel nitelikleri ve yetenekleri hakkında konuşmaları beklenmektedir. Bu durumda, öğrencilerinizin konuşma becerilerini değerlendirmeniz gerekiyor.
 - Öğrencilerinizin konuşma becerilerini nasıl değerlendirirsiniz?
 - Dil bileşenlerini nasıl değerlendirirsiniz?
4. Bir ilkokulda İngilizce öğretmeni olarak çalışıyorsunuz. Öğrencilerinizin (4. Sınıf/A1 Seviyesi) sevilen/sevilmeyen aktiviteler ve günlük rutinler veya zaman ifadeleri hakkındaki bilgileri anlamaları beklenmektedir. Bu durumda, öğrencilerinizin dinleme becerilerini değerlendirmek istiyorsunuz.
 - a) Öğrencilerinizin dinleme becerilerini nasıl değerlendirirsiniz?
 - b) Dil bileşenlerini nasıl değerlendirirsiniz?
 - c) Değerlendirmeleri nasıl puanlarsınız?
 - d) Sonuçları nasıl yorumlar ve kullanırsınız?

5. İlkokulda çalışan bir İngilizce öğretmenisiniz. Öğrencilerinizden (4. Sınıf/A1 Seviyesi), kısa ve basit ifadelerle izin isteyip/vererek ve ricada bulunarak diğer insanlarla etkileşim kurmaları beklenmektedir. Bu durumda, öğrencilerinizin konuşma becerilerini değerlendirmek istiyorsunuz.
- a) Öğrencilerinizin konuşma becerilerini nasıl değerlendirirsiniz?
 - b) Dil bileşenlerini nasıl değerlendirirsiniz?
 - c) Değerlendirmeleri nasıl puanlarsınız?
 - d) Sonuçları nasıl yorumlar ve kullanırsınız?
6. Ergen yaştaki öğrencilere (A2 Seviyesi) İngilizce öğretiyorsunuz ve öğrencilerinizden, kısa ve basit sözlü metinlerde macera ve turizmle ilgili ana noktaları ve ayrıntılı bilgileri belirlemeleri bekleniyor. Bu durumda, öğrencilerinizin dinleme becerilerini değerlendirmek istiyorsunuz.
- a) Öğrencilerinizin dinleme becerilerini nasıl değerlendirirsiniz?
 - b) Dil bileşenlerini nasıl değerlendirirsiniz?
 - c) Değerlendirmeleri nasıl puanlarsınız?
 - d) Sonuçları nasıl yorumlar ve kullanırsınız?
7. Sınıfınız ergen yaştaki öğrencilerden (A2+/B1 Seviyeleri) oluşuyor ve onlardan bir hikâyeyi yeniden anlatmaları veya canlandırmaları ve geçmiş alışkanlıklar veya gelenekler hakkında konuşmaları bekleniyor. Bu durumda, öğrencilerinizin konuşma becerilerini değerlendirmek istiyorsunuz.
- a) Öğrencilerinizin konuşma becerilerini nasıl değerlendirirsiniz?
 - b) Dil bileşenlerini nasıl değerlendirirsiniz?
 - c) Değerlendirmeleri nasıl puanlarsınız?
 - d) Sonuçları nasıl yorumlar ve kullanırsınız?
8. Ergen yaştaki öğrencilerinizden (B1+/B2 Seviyeleri), özetlemek için geçmiş alışkanlıklar ve deneyimlerle ilgili bir metni analiz etmeleri ve ünlü kişilerin biyografilerindeki olayları sıralamaları bekleniyor. Bu durumda, öğrencilerinizin okuma becerilerini değerlendirmek istiyorsunuz.
- a) Öğrencilerinizin okuma becerilerini nasıl değerlendirirsiniz?
 - b) Dil bileşenlerini nasıl değerlendirirsiniz?
 - c) Değerlendirmeleri nasıl puanlarsınız?
 - d) Sonuçları nasıl yorumlar ve kullanırsınız?

9. Lisede okuyan öğrencilerinizden (B2+ Seviyesi) haberleri, hikâyeleri veya geçmiş olayları kendi ifadeleriyle anlatmaları ve bir e-posta veya şikâyet mektubu yazmaları bekleniyor. Bu durumda, öğrencilerinizin yazma becerilerini değerlendirmek istiyorsunuz.

- a) Öğrencilerinizin yazma becerilerini nasıl değerlendirirsiniz?
- b) Dil bileşenlerini nasıl değerlendirirsiniz?
- c) Değerlendirmeleri nasıl puanlarsınız?
- d) Sonuçları nasıl yorumlar ve kullanırsınız?

10. Sınıf/Seviye: 9/A2

Tema: Çevrem

Öğrenme Çıktıları (MEB, 2018, s. 24)

Dinleme

Öğrenciler kaydedilmiş bir metin/videonun konusu ile ilgili soruları cevaplayabileceklerdir.

Öğrenciler yönergelerle ilgili bilgileri dinlerken nesnelerin yerlerini bulabileceklerdir.

Telaffuz

Öğrenciler /i/ ve /i:/ seslerini ayırt edebilecektir. Örneğin, sit /sıt/ seat /si:t/

Konuşma

Öğrenciler mahalleleri hakkında soru sorabilecekler ve mahallerini tasvir edebileceklerdir.

Öğrenciler çevrelerindeki insanları, yerleri ve nesneleri karşılaştırabileceklerdir.

Öğrenciler nesnelerin ve yerlerin konumu hakkında sorular sorabilecek ve sorulan soruları cevaplayabileceklerdir.

Okuma

Öğrenciler mahalleleri/şehirleri vb. hakkında belirli bilgiler için basit bir metni okuyabileceklerdir.

Yazma

Öğrenciler, farklı ülkelerdeki/Türkiye'deki şehirleri karşılaştıran bir tablo doldurabileceklerdir.

Öğrenciler farklı ortamları basit cümleler ve ifadelerle betimleyebileceklerdir.

Dersinizde yukarıdaki üniteyi işlediniz ve daha sonra, yukarıda verilen konuyu ve öğrenme çıktılarını kullanarak öğrencilerinizi dil becerileri ve dil bileşenleri açısından

değerlendirmek istiyorsunuz. Bu durumda, öğrencilerinizi dil becerileri ve dil bileşenleri açısından nasıl değerlendirirsiniz?

1. Tümüleşik değerlendirme (Nasıl değerlendireceğinizi ayrıntılı olarak açıklayınız.)
Ya da
 2. Beceriler birbirinden bağımsız değerlendirme (Nasıl değerlendireceğinizi ayrıntılı olarak açıklayınız.)
11. Sınıfınızda görme azlığı yaşayan bir öğrenciniz var ve bu nedenle, öğrenciniz yazılı materyalleri yeterince iyi göremiyor. Ancak, öğrencilerinizin dil becerilerini değerlendirmek istiyorsunuz ve değerlendirmelerinizin her öğrenci için adil olması gerektiğinin farkındasınız.
- Alıcı ve üretici becerilere ilişkin değerlendirmelerinizde adaleti sağlamak için neler yaparsınız?
12. Farklı millet ve kültürlerden (örn. Türk, Kürt, Arap vb.) öğrencilerden oluşan bir sınıfta İngilizce öğretiyorsunuz ve öğrencilerinizin İngilizce dil becerilerini değerlendirmek istiyorsunuz. Ancak, değerlendirmelerinizin tüm öğrenciler için adil olması gerektiğinin farkındasınız.
- Alıcı ve üretici becerilere ilişkin değerlendirmelerinizde adaleti sağlamak için neler yaparsınız?
13. Lise son sınıfta okuyan öğrencilere İngilizce öğreten bir öğretmensiniz. Akademik yılın sonunda öğrencileriniz üniversiteye girebilmek için önemli sonuçları olan bir yabancı dil sınavına gireceklerdir. Ancak, bu çoktan seçmeli sınavda sadece okuma becerileri ile kelime ve gramer bilgisi değerlendirilmektedir. Bu durumda, öğrencilerinizi İngilizce dil becerileri ve dil bileşenleri açısından derslerinizde nasıl değerlendirirsiniz?

Appendix 13. Scenarios for Foreign Language Assessment (English Version)

The aim of the Scenarios for Foreign Language Assessment (SfFLA) is to gather information about how you will be able to assess your students when faced with the given situations. This tool involves 13 distinct scenarios and 33 questions in total. Please answer each question in detail as much as possible by taking into consideration each of the scenarios given below.

1. Suppose that you have a new class consisting of English as a foreign language learners. However, you do not know your students' English levels and you have to learn about them before beginning to teach these students. How do you assess your students in order to detect their strengths and weaknesses in English?
2. You have a class consisting of young learners at A1 level. These students are expected to (a) identify the names and colors of classroom objects and (b) follow short and simple oral directions and instructions in English. In this case, you want to assess your students' listening skills.
 - a) How do you assess your students' listening skills?
 - b) How do you assess language components (vocabulary and grammar)?
3. You are teaching English to young learners (A1 level) at a primary school. Your students are expected to introduce people in their families and talk about the physical qualities and abilities of those people by producing short and simple sentences in English. In this case, you have to assess the speaking skills of your students.
 - a) How do you assess your students' speaking skills?
 - b) How do you assess language components?
4. You are working as an English teacher at a primary school. Your students (4th grade/A1 level) are expected to understand the information about likes/dislikes and daily routines or time expressions. In this case, you want to assess your students' listening skills.
 - a) How do you assess your students' listening skills?
 - b) How do you assess language components?
 - c) How do you score the assessments?
 - d) How do you interpret the results and use them?
5. You are an English teacher at a primary school. Your students (4th grade/A1 level) are expected to interact with other people by asking for or giving permission and making requests by producing short and simple utterances. In this case, you want to assess your students' speaking skills.
 - a) How do you assess your students' speaking skills?
 - b) How do you assess language components?
 - c) How do you score the assessments?
 - d) How do you interpret the results and use them?
6. You are teaching English to adolescents and your students (A2 level) are expected to identify the main points and the specific information about adventures and tourism in the short and simple oral texts in English. In this case, you want to assess your students' listening skills.
 - a) How do you assess your students' listening skills?

- b) How do you assess language components?
 - c) How do you score the assessments?
 - d) How do you interpret the results and use them?
7. Your class consists of adolescents (A2+/B1 levels) and they are expected to retell or act out a story and talk about the past habits or traditions. In this case, you want to assess your students' speaking skills.
- a) How do you assess your students' speaking skills?
 - b) How do you assess language components?
 - c) How do you score the assessments?
 - d) How do you interpret the results and use them?
8. Your adolescent students (B1+/B2 levels) are expected to analyze a text about past habits and experiences in order to summarize it and order the events in the biography of famous people. In this case, you want to assess your students' reading skills.
- a) How do you assess your students' reading skills?
 - b) How do you assess language components?
 - c) How do you score the assessments?
 - d) How do you interpret the results and use them?
9. Your students (B2+ level) who are studying at a high school are expected to paraphrase news, stories, or past-events and write an email or a letter of complaint. In this case, you want to assess these students' writing skills.
- a) How do you assess your students' writing skills?
 - b) How do you assess language components?
 - c) How do you score the assessments?
 - d) How do you interpret the results and use them?

10. Grade/Level: 9/A2

Theme: My Environment

Learning Outcomes (MONE, 2018, p. 24):

Listening

Students will be able to respond to the questions related to the topic of a recorded text/video.

Students will be able to locate the things as they listen to information about the instructions.

Pronunciation

Students will be able to differentiate /ɪ/ and /i:/ sounds. Eg. sit /sɪt/ seat /si:t/

Speaking

Students will be able to ask about and describe their neighbourhood.

Students will be able to compare people, places and objects around them.

Students will be able to ask and answer questions about location of things and places.

Reading

Students will be able to read a simple text for specific information about their neighbourhood/city etc.

Writing

Students will be able to fill in a chart comparing cities in different countries/Turkey.

Students will be able to describe different environments in simple sentences and phrases.

In your lesson, you have covered the unit above and then, you want to assess your students in terms of language skills and language related components (vocabulary and grammar, and pronunciation) by drawing on the topic and the learning outcomes given above. In this case, how do you assess your students in terms of the language skills and language related components?

1. Integrated assessment (Please explain in detail how you will assess.)

Or

2. Discrete assessment (Please explain in detail how you will assess.)

11. In your class, there is a student with low vision and therefore, s/he has difficulty in seeing the written materials well enough. However, you want to assess all your students' language skills and you are aware that your assessments should be fair for every student.
 - What do you do to ensure fairness in your assessments of receptive and productive skills?
12. You are teaching English in a class consisting of the students from different nations and cultures (e.g. Turk, Kurd, Arab, etc.) and you want to assess your students' language skills. However, you are aware that your assessments should be fair for all students.
 - What do you do to ensure fairness in your assessments of receptive and productive skills?
13. You are a teacher teaching English to senior high school students. At the end of the academic year your students are going to take a high-stakes foreign language exam in order to enter university. However, only reading skills and knowledge of vocabulary and grammar are assessed in this multiple-choice exam. In this case, how do you assess your students in terms of English language skills and language components in your lessons?

Appendix 14. Ethics Committee Approval

Evrak Tarih ve Sayısı: 26.10.2020-E.115003



T.C.
GAZİ ÜNİVERSİTESİ
Ölçme Değerlendirme Etik Alt Çalışma Grubu



Sayı : 91610558-302.08.01-
Konu : Bilimsel ve Eğitim Amaçlı

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 19.08.2020 tarih ve E.86716 sayılı yazı

İlgi yazınız ile göndermiş olduğunuz, Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı, İngiliz Dili Eğitimi Bilim Dalı **Doktora Öğrencisi Murat ŞÜKÜR'un, Prof.Dr.Gonca YANGIN EKŞİ'nin** danışmanlığında yürüttüğü **"Hizmet Öncesi İngilizce Öğretmenlerinin Yabancı Dilde Ölçme Değerlendirme Okuryazarlığı Geliştirme Sürecinin Betimlenmesi"** adlı tez çalışması ile ilgili konu Kurulumuzun **08.09.2020** tarih ve **08** sayılı toplantısında görüşülmüş olup,

İlgilinin çalışmasının, yapılması planlanan yerlerden izin alınması koşuluyla yapılmasında etik açıdan bir sakınca bulunmadığına oybirliği ile karar verilmiş ve karara ilişkin imza listesi ekte gönderilmiştir.

Bilgilerinizi ve gereğini rica ederim.

e-imzalıdır
Prof. Dr. İsmail KARAKAYA
Kurul Başkanı

Araştırma Kod No: 2020 -562

Ek: 1 Liste



Emniyet Mahallesi Bandırma Caddesi No :6/1 06560 Yenimahalle/ANKARA
Tel:0 (312) 202 20 57 - 0 (312) 2... Faks:0 (312) 202 38 76
İnternet Adresi :http://etikkomisyon.gazi.edu.tr/

Bilgi için :Nursel Güner
Birim Evrak Sorumlusu
Telefon No:202 20 57

Bu belge 5070 sayılı Elektronik İmza Kanununun 5. Maddesi gereğince güvenli elektronik imza ile imzalanmıştır.



GAZİLİ OLMAK AYRICALIKTIR..