

**REPUBLIC OF TÜRKİYE
ÇUKUROVA UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LANGUAGE TEACHING**

**DEVELOPMENT OF EFL INSTRUCTORS' COGNITIONS ABOUT
LANGUAGE ASSESSMENT LITERACY THROUGH
REFLECTIVE READING AND DISCUSSION**

Yaşar Üstün KAPLAN

PhD. DISSERTATION

ADANA / 2022

REPUBLIC OF TÜRKİYE
ÇUKUROVA UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LANGUAGE TEACHING

**DEVELOPMENT OF EFL INSTRUCTORS' COGNITIONS ABOUT
LANGUAGE ASSESSMENT LITERACY THROUGH
REFLECTIVE READING AND DISCUSSION**

Yaşar Üstün KAPLAN

Supervisor: Assoc. Prof. Dr. Gülden İLİN

Member of Examining Committee: Prof. Dr. Neşe CABAROĞLU

Member of Examining Committee: Prof. Dr. Mahmut Oğuz KUTLU

Member of Examining Committee: Assoc. Prof. Dr. Fadime YALÇIN ARSLAN

Member of Examining Committee: Assist. Prof. Dr. Hasan Çağlar BAŞOL

PhD. DISSERTATION

ADANA / 2022

To Çukurova University Institute of Social Sciences,

We certify that this dissertation is satisfactory for the award of the degree
DOCTOR of PHILOSOPHY in the Department of English Language Teaching

Supervisor: Assoc. Prof. Dr. Gülден İLİN

Member of Examining Committee: Prof. Dr. Neşe CABAROĞLU

Member of Examining Committee: Prof. Dr. Mahmut Oğuz KUTLU

Member of Examining Committee: Assoc. Prof. Dr. Fadime YALÇIN ARSLAN

Member of Examining Committee: Assist. Prof. Dr. Hasan Çağlar BAŞOL

I certify that this dissertation confirms the formal standards of the Institute of Social
Sciences. / / 2022

Prof. Dr. Serap ÇABUK

Enstitü Müdürü

Note: The uncited usage of the reports, charts, figures, and photographs in this thesis, whether original or quoted from other sources, is subject to the law of Works of Art and Thought No: 5846.

NOT: Bu tezde kullanılan ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunu'ndaki hükümlere tabidir.

ETİK BEYANI

Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde ve ortaya çıkan sonuçlarda herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim. / / 2022

Yaşar Üstün KAPLAN

DEDICATION



To my precious daughters

Nursema KAPLAN

Bilge KAPLAN

“Sadece ikimizin uyandıđı saatlerde duruyor zaman çünkü sadece sen tutuklarsın beni apansız uyanış gibi. Gel kızım sokul bana bir kez daha alayım kokusunu benim küçük bahçemin. Büyüsen de, gitsen de hala bekliyor gibi beni uzanmış küçük ellerin”.

ABSTRACT**DEVELOPMENT OF EFL INSTRUCTORS' COGNITIONS ABOUT LANGUAGE
ASSESSMENT LITERACY THROUGH REFLECTIVE READING AND
DISCUSSION****Yaşar Üstün KAPLAN****PhD. Dissertation, Department of English Language Teaching****Supervisor: Assoc. Prof. Dr. Gülden İLİN****December 2022, 319 pages**

Research on teacher cognition in L2 education has received increased attention in the past two decades. Although assessment in instruction is frequently seen as a way to assess student achievement (Çelik & Türkan, 2020), it is a crucial component of all stages of teaching and learning (Çelik & Coombe, 2021). This current study concentrated on shedding light on the nature of language instructors' language assessment cognitions. It also intended to identify participant teachers' strengths and weaknesses concerning language assessment and design reflective reading and discussion (RRD) sessions depending on teachers' needs and expectations. In addition, exploring probable changes in participants' cognitions of language assessment as a result of RRD sessions was another objective of the study. To this end, four EFL instructors working at a foreign language school participated in this research. The study was separated into three phases: pre-RRD sessions, during RRD sessions, and post-RRD sessions. During the pre-RRD session, participants were interviewed and observed to identify their beliefs and practices regarding language assessment. Also, their strengths and weaknesses concerning with the issue were explored in the initial part of the study. Next, teachers were involved in a set of RRD sessions. During those sessions, participants read the reading materials and discussed the target topic of each specific session. They also filled in KWL charts (What I know, What I want to know and What I have learnt) and wrote reflection reports after each RRD session. As for the third part of the study, developments in teachers' language assessment-related cognitions, if any, were figured out. The study was designed as a case study based on the qualitative approach. Theoretical framework of the study was based on Vygotsky (1978)'s socio-cultural theory. The data were collected through multiple sources

such as pre and follow-up semi-structured interviews, classroom observations, KWL charts, and reflection reports. Inductive content analysis was employed to analyse the obtained data. The results showed that there is a development in the cognitions of participants considering language assessment after the RRD sessions. In the light of emerging results, discussions were provided to examine the development of participants' assessment-related cognitions following RRD sessions. Moreover, recommendations about teachers' pre-service and in-service education regarding language assessment have been made on the basis of findings and the discussions.

Keywords: Teacher cognition, assessment, language assessment literacy, development, in-service English language teacher, reflection



ÖZET

YANSITICI OKUMA VE TARTIŞMA YOLUYLA İNGİLİZCE OKUTMANLARININ DİLDE ÖLÇME DEĞERLENDİRME OKURYAZARLIĞI BİLİŞLERİNİN GELİŞİMİ

Yaşar Üstün KAPLAN

Doktora Tezi, İngiliz Dili Eğitimi Ana Bilim Dalı

Danışman: Doç. Dr. Gülден İLİN

Aralık 2022, 319 sayfa

İkinci dil eğitiminde öğretmen bilişi üzerine yapılan araştırmalar son yirmi yılda artan bir ilgi görmüştür. Ölçme değerlendirme sıklıkla öğrenci başarısını değerlendirmenin bir yolu olarak görülse de (Çelik ve Türkan, 2020), o, öğretim ve öğrenmenin tüm aşamalarının önemli bir bileşenidir (Çelik ve Coombe, 2021). Bu mevcut çalışma, İngilizce öğretim görevlilerinin var olan dilde değerlendirme bilişlerine ışık tutmaya odaklanmaktadır. Ayrıca, bu çalışmada katılımcıların dilde değerlendirme ile ilgili güçlü ve zayıf yönlerini belirlenmesi ve onların değerlendirme ihtiyaçlarına göre yansıtıcı okuma ve tartışma oturumlarının tasarlanması amaçlanmaktadır. Ek olarak, eğer varsa, yansıtıcı okuma ve tartışma oturumlarının sonucunda katılımcı öğretmenlerin dilde değerlendirme okuryazarlığı bilişlerindeki gelişimi keşfetmek de çalışmanın amaçları arasında yer almaktadır. Bu amaçla, bu araştırmaya bir yabancı diller yüksek okulunda görev yapan dört İngilizce okutmanı katıldı. Çalışma, yansıtıcı okuma ve tartışma oturumları öncesi, yansıtıcı okuma ve tartışma oturumları sonrası ve yansıtıcı okuma ve tartışma oturumları sonrası olmak üzere üç kısma ayrıldı. Çalışmanın yansıtıcı okuma ve tartışma oturumları öncesi aşaması sırasında, katılımcılarla onların değerlendirmeye özellikle dil değerlendirmesine ilişkin inanç ve uygulamalarını belirlemek için yarı-yapılandırılmış görüşme yapıldı ve sonra öğretmenler sınıflarında gözlemlendi. Ayrıca öğretmenlerin dil değerlendirmesindeki güçlü ve zayıf yönleri de bu bölümde açığa çıkarıldı. Daha sonra onlar, ihtiyaçları ve beklentileri göz önünde bulundurularak tasarlanmış bir dizi yansıtıcı okuma ve tartışma oturumuna katıldı. Bu oturumlar sırasında katılımcılar, her oturumun hedef konusu hakkında önceden kendilerine verilen okuma metinlerini okudu ve sonra hedef konuyu birlikte tartıştı. Araştırmanın üçüncü bölümünde ise öğretmenlerin katıldıkları yansıtıcı okuma ve tartışma seansları ardından dilde

ölçmeye ilişkin bilişlerinde olası gelişim ortaya çıkarılmaya çalışıldı. Çalışma, nitel yaklaşıma dayalı bir durum çalışması olarak tasarlandı ve çalışmanın teorik çerçevesini Vygotsky (1978)'nin sosyo-kültürel teorisi oluşturdu. Veriler, bir devlet üniversitesindeki yabancı diller yüksekokulunda çalışan okutmanlarla yarı yapılandırılmış görüşmeler, sınıf gözlemi, KWL şeması ve yansıtma raporları gibi çoklu kaynaklar aracılığıyla toplandı. Toplanan verilerin analizi için tümevarımsal içerik analizi kullanıldı. Ortaya çıkan sonuçlar katılımcıların RRD seansları sonrasında dilde değerlendirmeye yönelik bilişlerinde bir gelişimin olduğunu göstermektedir. Ayrıca sonuçlara dayanarak, öğretmenlerin dilde değerlendirmeyle ilgili bilişlerinin gelişimini incelemek için yorumlar yapıldı. Elde edilen bulgular ve tartışmalara dayalı olarak dilde değerlendirme yönelik öğretmenlerin hizmet öncesi ve hizmet içi eğitime yönelik önerilerde bulunuldu.

Anahtar Sözcükler: Öğretmen bilişi, değerlendirme, dilde ölçme değerlendirme okuryazarlığı, gelişim, hizmet içi İngilizce öğretmeni, yansıtma

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my supervisor, Assoc. Prof. Dr. Glden İLİN for her invaluable advice, continuous support, tolerance and patience during my PhD journey. She has been more than a supervisor encouraging me all the time. Without her help, contribution and guidance, it would not have been possible to complete this dissertation. I will be grateful to my supervisor throughout my life. I would also like to thank Çetin İLİN. My dear teachers made me feel like I was their son. I am grateful to them.

In addition, I would like to thank Prof Dr. Neşę CABAROĞLU, who has shed light on my way since my undergraduate education years, for her contributions that I cannot express in words. I feel very lucky to have been her student for 13 years. I must emphasise that I will remain as her "le Papillion" that I portrayed while taking Teaching Language Skills course given by her in 2010.

I want to express my profound gratitude to Prof. Dr Mahmut Oğuz KUTLU for his invaluable support and insightful comments on my dissertation.

I gratefully appreciate the outstanding contributions of Assoc. Prof. Dr. Fadime YALÇIN ARSLAN with her constructive feedback as a jury member of the doctoral defence committee. My thanks also go to Assist Prof. Dr Hasan Çağlar BAŞOL for agreeing to participate in the defence jury with his precious feedbacks.

I would also love to express my sincere thanks to my colleagues who participated in this study. Without their enthusiastic participation and commitment, the study could not have attained its full potential.

I want to thank my dearest friend Kismet, who has always devoted himself to my professional development since our high school years. I am grateful to him for being with me in my most difficult time since my childhood, for constantly encouraging me to reach this day.

I would also like to express my endless thanks to my esteemed family, who believe in me at every stage of my life. First, I am grateful especially to my grandmother Ayşę BOZDOĞANOĞLU for everything that she taught me. I learned not to give up from her! Secondly, I would especially like to express my deepest love and gratitude to my aunt, my second mother, Selma BOZDOĞANOĞLU, for her existence, for the endless love she has shown me, and for protecting and supporting me under all circumstances. I will always love you my dear aunty. Next, million thanks go to Perihan KAPLAN. She always get me to feel precious. I also learned to be stubborn from her.

I would also like to thank my dear mother, Nursema KAPLAN. Actually, this success is primarily thanks to her support. In her black eyes, I saw both sadness and hope and that's why I didn't give up and I won't... Actually my every success dedicates to her! As for my dear brave brother Battal Servet KAPLAN, even if he will be thirty years old and the father of a sweet and naughty twin, he will always remain as my 3-year-old brother. I love him more than anything in my life. We overcame all the problems together. To our father, Hayati KAPLAN, who passed away 26 years ago, we know he is with us. Rest in peace.

To my dear wife Şengül KAPLAN... I am grateful to her for the sacrifices she made. I would not complete my graduate education without her help, support and especially her patience. I have been a graduate student since the day when we first met. From now on, I promise that I will compensate for everything that we postponed. *“Dance me to the end of love!”*

To my little daughters. I cannot find suitable words for describing their value in my life. Sema and Bilge, for the strength they gave me, for the love they gave, for the feelings they make me feel, I thank them million times. I would not have accomplished the person I am today without their existence. The only thing I wish with all my heart is for them to grow into intellectual, strong, self-confident, successful, free and happy women. *“Victory is for those who can say "Victory is mine". Success is for those who can begin saying "I will succeed" and say "I have succeeded" in the end”.* Mustafa Kemal ATATÜRK. My dear daughters, believe in yourself and never give up. Then, success will be with you.

As for Çukurova University, I would like to thank our magnificent campus, R1 classrooms, and the Dean's Office of the Faculty of Education for hosting me in the best 13 years of my life. Even though I will work at different universities, it will remain the most precious!

The study has also been supported by Scientific Research Projects of Çukurova University, with the project code SDK-2020-12848

Yaşar Üstün KAPLAN

Adana / 2022

TABLE OF CONTENTS

ABSTRACT	iv
ÖZET	vi
ACKNOWLEDGEMENTS	viii
TABLE OF CONTENTS	x
ABBREVIATIONS	xvii
LIST OF TABLES	xviii
LIST OF FIGURES	xxii
LIST OF APPENDICES	xxiii
CHAPTER I	1
INTRODUCTION	1
1.1. Background to Study	1
1.2. Statement of the Problem	3
1.3. Significance of the Study	5
1.4. Purpose of the Study	7
1.5. Operational Definitions	8
CHAPTER II	9
LITERATURE REVIEW	9
2.1. Language Teacher Cognition	9
2.2. Dimensions of Language Teacher Cognition	11
2.3. Factors that Shape Language Teacher Cognition	13
2.4. Relationship between Language Teacher Cognition and Classroom Practices	15
2.5. Research in Teacher Cognition	16
2.6. Research on Language Teacher Cognition	17
2.7. Studies on EFL Teachers' Cognitions in the Turkish Context	22
2.8. Assessment in Language Teaching	23
2.9. Basic Concepts in Assessment: Assessment, Measurement, and Evaluation	23
2.10. Types of Assessment	24
2.11. Principles of Language Assessment	26
2.12. Assessment Literacy	29
2.13. Language Assessment Literacy	31
2.14. Theoretical Framework	35
CHAPTER III	39

MEHTOD.....	39
3.1. Introduction	39
3.2. The Research Paradigm and the Research Design	39
3.3. Participants and the Sampling Method.....	45
3.4. The Context of the Study	47
3.5. The Research Procedure.....	48
3.5.1. The Reflective Reading and Discussion Procedure	54
3.6. Data Collection Tools.....	61
3.6.1. Semi-structured interview	61
3.6.2. Classroom Observation	62
3.6.3. KWL Chart.....	63
3.6.4. Reflection Reports.....	64
3.7. Data Analysis	65
3.7.1. Transcription	68
3.7.2. Coding	69
3.8. Ethical Considerations.....	72
3.8.1. Informed Consent.....	72
3.8.2. Anonymity and Confidentiality.....	73
3.8.3. Data Security	74
3.8.4. Honesty and Trust	75
3.9. Trustworthiness of the Study.....	76
3.10. Researcher Role.....	80
CHAPTER IV	82
FINDINGS	82
4.1. Introduction	82
4.2. What is EFL Instructors' Nature of Cognitions Regarding Language Assessment?	83
4.2.1. Teachers' Assessment Perception	84
4.2.1.1. Teachers' Definition of Assessment	85
4.2.1.2. Teachers' Views on Classroom Assessment.....	86
4.2.1.3. Teachers' Purposes of Assessment	87
4.2.1.4 Teachers' Views of Ideal Time for Assessment.....	90
4.2.1.5. Teachers' Views on Effective Assessment	92
4.2.1.6. Teachers' Initial Perception of Assessment and Change in it over Time	94
4.2.2. Teachers' Views about Assessment Technique	97

4.2.2.1 Teachers' Assessment Technique Knowledge and Preference	97
4.2.2.2 Teachers' Purpose of Using Assessment Technique.....	100
4.2.2.3. Reasons Behind Teachers' Preference for Assessment Technique.....	101
4.2.2.4. Teachers' Perception of Effective Assessment Technique and Its Characteristics...	103
4.2.2.5. Teachers' Perception of Alternative Assessment.....	105
4.2.3. Teachers' Views Concerning Assessment Results.....	107
4.2.3.1 Teachers' Opinions of Assessment Results	107
4.2.3.2 Teachers' Purposes for Using Assessment Results.....	108
4.2.4. Teachers' Views Regarding the Influence of Institutions on Assessment.....	110
4.2.4.1 Teachers' Opinions about Institutional Expectations Regarding Assessment.....	111
4.2.4.2. Institutional Expectations of Assessment.....	113
4.2.4.3. Influence of institution on Teachers' Assessment Perception and Practices	114
4.2.5. Teachers' Thoughts on the Influence of Educational Policy of the Country on Assessment.....	116
4.2.5.1. Expectations of Assessment Policy from Teachers.....	117
4.2.5.2. Influence of Educational Policy on Teachers' Assessment Perceptions and Practices	118
4.3. What are the Participants' Strengths and Weaknesses Considering Language Assessment as They Perceive?.....	120
4.3.1. <i>Teachers'</i> Strengths Related to Assessment.....	120
4.3.1.1. Teachers' Strengths Concerning Assessment Results.....	122
4.3.1.1.1. Teachers' Other Strengths in Assessment Results	122
4.3.1.1.2. Teachers' Strengths in Assessment Results Usage	124
4.3.1.1.3. Teachers' Strengths in Assessment Data Evaluation	125
4.3.1.2. Teachers' Assessment Techniques Related Strengths.....	127
4.3.1.2.1. Teachers' Strengths in Assessment Technique Knowledge and Implementation..	127
4.3.1.2.2. Teachers' Strengths in Considerations for Selection of Techniques.....	129
4.3.1.2.3. Teachers' Strengths in Purpose for Assessment Technique Use	131
4.3.1.3. Teachers' Strengths in Developing and Administering Exams	132
4.3.1.3.1. Teachers' Strengths in Test Preparation.....	132
4.3.1.3.2. Teachers' Strengths in Administering Exams	134
4.3.1.3.3. Teachers' Test Item Related Strengths.....	136
4.3.1.4. Teachers Strengths About Introductory Topics in Assessment	137
4.3.1.4.1. Teachers' Strengths in Assessment Related Definitions.....	137

4.3.1.4.2. Teachers' Strengths in Purpose of Assessment.....	139
4.3.1.5. Teachers' Strengths in Ethical Considerations in Assessment.....	140
4.3.1.5.1. Teachers' Strengths Regarding Views of Ethics.....	140
4.3.1.5.2. Teachers' Strengths in Ethical Practices	142
4.3.1.6. Teachers' Strengths in Scoring and Grading.....	143
4.3.1.6.1. Teachers' Strengths in Grading.....	143
4.3.1.6.2. Teachers' Strengths in Scoring	146
4.3.1.7. Teachers' Other Strengths Regarding Assessment	147
4.3.1.7.1. Teachers' Other Assessment Related Strengths.....	147
4.3.1.7.2. Teachers' Strengths in Avoiding from Measurement Error.....	148
4.3.2. <i>Teachers' Weaknesses Related to the Assessment</i>	150
4.3.2.1. Teachers' Weaknesses Related to Test Construction and Test Items	152
4.3.2.1.1. Teachers' Weaknesses in Test Construction	152
4.3.2.1.2. Teachers' Weaknesses in Item Analysis	155
4.3.2.1.3. Teachers' Weaknesses in Including Test Items in a Test.....	155
4.3.2.1.4. Teachers' Weaknesses in Item Writing.....	157
4.3.2.1.5. Teachers' Weaknesses in the Purpose of Item Usage	159
4.3.2.2. Teacher's Weaknesses about Alternative Assessment.....	160
4.3.2.2.1. Teachers' Weaknesses in Knowledge of Alternative Techniques	160
4.3.2.2.2. Teachers' Weaknesses in Performance Assessment and Rubric.....	163
4.3.2.2.3. Teachers' Weaknesses in Practicing Alternative Assessment Techniques	164
4.3.2.2.4. Teachers' Weaknesses in Alternative Assessment.....	165
4.3.2.3. Teachers' Weaknesses in Introduction Topics in Assessment.....	166
4.3.2.3.1. Teachers' Weaknesses in Assessment-related Definitions	167
4.3.2.3.2. Teachers' Weaknesses in Assessment Types.....	168
4.3.2.3.3. Teachers' Weaknesses in Purposes of Assessment.....	169
4.3.2.4. Teachers' Weaknesses about Assessment Results	171
4.3.2.4.1. Teachers' Weaknesses in Interpreting Assessment Results.....	171
4.3.2.4.2. Teachers' Other Weaknesses Related to Assessment Results	172
4.3.2.4.3. Teachers' Weaknesses in Using Assessment Results	174
4.3.2.5. Teachers' Weaknesses Related to the Assessment Technique.....	176
4.3.2.5.1. Teachers' Weaknesses in Knowledge and Implementation of Assessment Technique	176
4.3.2.5.2. Teachers' Weaknesses in Criteria for Technique Selection.....	178

4.3.2.5.3. Teachers' Weaknesses in the Purpose of Technique Usage.....	180
4.3.2.6. Teachers' Weaknesses in Grading and Scoring	180
4.3.2.6.1. Teachers' Weaknesses in Grading	181
4.3.2.6.2. Teachers' Weaknesses in Scoring	182
4.3.2.7. Teachers' Additional Weaknesses in Assessment	183
4.3.2.7.1. Teachers' Weaknesses about Ethics in Assessment.....	183
4.3.2.7.2. Teachers' Other Assessment Related Weaknesses	185
4.3.2.8. Teachers' Weaknesses Regarding Assessment Principles	186
4.3.2.8.1. Teachers' Weaknesses in Validity	186
4.3.2.8.2. Teachers' Weaknesses in Reliability.....	187
4.3.2.8.3. Teachers' Weaknesses in Influence of Assessment	189
4.4. What is the Perceived Development of EFL Instructors Considering Language Assessment, if any?	189
4.4.1. Teachers' Developments Concerning Assessment Techniques	192
4.4.1.1. Teachers' Developments Regarding Performance Assessment and Rubric	192
4.4.1.2. Teachers' Developments in Knowledge of Alternatives.....	194
4.4.1.3. Teachers' Developments in Specific Alternative Techniques	196
4.4.1.4. Teachers' Developments in What to Consider while Developing, Choosing or Adapting a Technique	199
4.4.1.5. Teachers' Developments about Assessment Technique	201
4.4.1.6. Teachers' Developments in Purpose of Using Assessment Techniques.....	202
4.4.2. Teachers' Developments in Writing Test Items and Test Construction	204
4.4.2.1. Teachers' Developments in Writing Kinds of Test Items.....	204
4.4.2.2. Teachers' Developments in Content of a Test in Test Preparation.....	206
4.4.2.3. Teachers' Developments in What to Consider while Writing Test Items.....	207
4.4.2.4. Teachers' Additional Developments Regarding Test Preparation	209
4.4.2.5. Teachers' Developments Considering Assessment Principles in Test Preparation ..	211
4.4.3. Teachers' Developments Regarding Assessment Results.....	212
4.4.3.1. Teachers' Developments in Use of Assessment Result	212
4.4.3.2. Teachers' Developments in Sharing Assessment Results.....	214
4.4.3.3. Teachers' Developments in Interpreting Assessment Results	216
4.4.3.4. Teachers' Developments in Making Decisions Based on Assessment Results	218
4.4.3.5. Teachers' Developments in Reflection of Assessment Result.....	220
4.4.4. Teachers' Additional Developments Regarding Assessment	221

4.4.4.1. Teachers' Developments in Ethics	222
4.4.4.2. Teachers' Developments in Administering Exams	224
4.4.4.3. Teachers' Developments in Characteristics for Quality Assessment.....	225
4.4.4.4. Teachers' Developments in What to Consider to Increase the Quality of Assessment	227
4.4.4.5. Teachers' Developments in Relationships in the Issues Regarding Assessment.....	230
4.4.5. Teachers' Developments in Introductory Issues about Assessment	231
4.4.5.1. Teachers' Developed Perceptions and Definitions of Assessment	231
4.4.5.2. Teachers' Developments Regarding Assessment Type	234
4.4.5.3. Teachers' Development of Assessment Purposes	235
4.4.6. Teachers' Developments in Assessment Principles	236
4.4.6.1. Teachers' Developments in Washback	236
4.4.6.2. Teachers' Developments in Validity.....	239
4.4.6.3. Teachers' Developments in Consistency	241
4.4.7. Teachers' Developments in Grading and Scoring.....	241
4.4.7.1. Teachers' Developments in Grading in General	242
4.4.7.2. Teachers' Additional Developments in Grading and Scoring	245
4.4.7.3. Teachers' Developments of Ethics in Grading	247
4.4.8. Teachers' Developments in Reliability	248
4.4.8.1. Teachers' Developments in Ensuring Reliability in Administering Exams.....	249
4.4.8.2. Teachers' Developments in Factors Influencing Reliability in Assessment.....	250
4.4.9. Teachers' Developments in Measurement Error.....	252
4.4.9.1. Teachers' Developments in Knowledge of Measurement Error	252
4.4.9.2. Teachers' Developments in How to Prevent Measurement Error.....	254
CHAPTER V	257
DISCUSSION	257
5.1. Discussion for the Research Question 1	257
5.2. Discussion for the Research Question 2.....	260
5.3. Discussion for the Research Question 3.....	261
CHAPTER VI	264
CONCLUSION	264
6.1. Introduction	264
6.2. Overview of the Study and Implications.....	264
6.3. Implications of the Study	270

6.4. Limitations of Study.....	273
6.5. Suggestions for Further Studies	273
6.6. What the Researcher has Learnt as a Result of Reflection and Discussion	275
REFERENCES.....	278
APPENDICES.....	300
CURRICULUM VITAE	319



ABBREVIATIONS

AL: Assessment Literacy

CLT: Communicative Language Teaching

EFL: English as a Foreign Language

EFL: English as a Foreign Language

ESL: English as a Second Language

LAL: Language Assessment Literacy

LTC: Language Teacher Cognition

RRD: Reflective Reading and Discussion

SLTE: Second Language Teacher Education



LIST OF TABLES

Table 1. Overall Research Procedure	49
Table 2. Content of Reflective Reading and Discussion Sessions	55
Table 3. Data Collection tools and their linkage to RQs	61
Table 4. Themes and Sub-themes Referring to Participants' Current Nature of Cognitions Regarding Assessment	83
Table 5. Teachers' Definition of Assessment.....	85
Table 6. Teachers' Purposes of Assessment.....	88
Table 7. Teachers' Views of Ideal Time for Assessment.....	90
Table 8. Teachers' Views on Effective Assessment.....	93
Table 9. Teachers' Initial Perception of Assessment and Change in it over Time.....	94
Table 10. Teachers' Assessment Technique Knowledge and Preference	98
Table 11. Teachers' Purpose of Using Assessment Technique.....	100
Table 12. Reasons Behind Teachers' Preference for Assessment Technique	102
Table 13. Teachers' Perception of Effective Assessment Technique and Its Characteristics	104
Table 14. Teachers' Perception of Alternative Assessment	105
Table 15. Teachers' Opinions of Assessment Results.....	107
Table 16. Teachers' Purposes for Using Assessment Results	108
Table 17. Teachers' Opinions about Institutional Expectations Regarding Assessment	111
Table 18. Institutional Expectations of Assessment	113
Table 19. Influence of Institution on Teachers' Assessment Perception and Practices	115
Table 20. Expectations of Assessment Policy from Teachers	117
Table 21. Influence of Educational Policy on Teachers' Assessment Perceptions and Practices.....	119
Table 22. Themes and sub-themes about strengths of the teachers concerning language assessment	120
Table 23. Teachers' Other Strengths in Assessment Results	122
Table 24. Teachers' Strengths in Assessment Result Usage	124
Table 25. Teachers' Strengths in Assessment Data Evaluation	126
Table 26. Teachers' Strengths in Assessment Technique Knowledge and Implementation.	128
Table 27. Teachers' Strengths in Considerations for Selection of Techniques.....	130
Table 28. Teachers' Strengths in Purpose for Assessment Technique Use.....	132
Table 29. Teachers' Strengths in Test Preparation.....	133
Table 30. Teachers' Strengths in Administering Exams	134

Table 31. Teachers' Test Item Related Strengths	136
Table 32. Teachers' Strengths in Assessment Related Definitions	137
Table 33. Teachers' Strengths in Assessment Purposes	139
Table 34. Teachers' Strengths Regarding Views of Ethics	141
Table 35. Teachers' Strengths in Ethical Practices	142
Table 36. Teachers' Strengths in Grading	144
Table 37. Teachers' Strengths in Scoring.....	146
Table 38. Teachers' Other Assessment Related Strengths	147
Table 39. Teachers' Strengths in Avoiding from Measurement Error	149
Table 40. Teachers' Weaknesses Related to the Assessment.....	150
Table 41. Teachers' Weaknesses in Test Construction	153
Table 42. Teachers' Weaknesses in Item Analysis	155
Table 43. Teachers' Weaknesses in Including Test Items in a Test.....	156
Table 44. Teachers' Weaknesses in Item Writing	158
Table 45. Teachers' Weaknesses in the Purpose of Item Usage	159
Table 46. Teachers' Weakness in Knowledge of Alternative Techniques.....	161
Table 47. Teachers' Weaknesses in Performance Assessment and Rubric.....	163
Table 48. Teachers' Weaknesses in Practicing Alternative Assessment Techniques	165
Table 49. Teachers' Weaknesses in Alternative Assessment.....	166
Table 50. Teachers' Weaknesses in Assessment-related Definitions.....	167
Table 51. Teachers' Weaknesses in Assessment Types	168
Table 52. Teachers' Weaknesses in Purposes of Assessment	170
Table 53. Teachers' Weaknesses in Interpreting Assessment Results	171
Table 54. Teachers' Other Weaknesses Related to Assessment Results.....	173
Table 55. Teachers' Weaknesses in Using Assessment Results	175
Table 56. Teachers' Weaknesses in Knowledge and Implementation of Assessment Technique	177
Table 57. Teachers' Weaknesses in Criteria for Technique Selection	178
Table 58. Teachers' Weaknesses in the Purpose of Technique Usage.....	180
Table 59. Teachers' Weakness in Grading.....	181
Table 60. Teachers' Weaknesses in Scoring	182
Table 61. Teachers' Weaknesses about Ethics in Assessment.....	184
Table 62. Teachers' Other Assessment Related Weaknesses.....	185
Table 63. Teachers' Weaknesses in Validity.....	186

Table 64. Teachers' Weaknesses in Reliability.....	188
Table 65. Teachers' Weaknesses in Influence of Assessment	189
Table 66. Teachers' Developments Considering Language Assessment	190
Table 67. Teachers' Developments Regarding Performance Assessment and Rubric.....	193
Table 68. Teachers' Developments in Knowledge of Alternatives.....	195
Table 69. Teachers' Developments in Specific Alternative Techniques.....	197
Table 70. Teachers' Developments in What to Consider while Developing, Choosing or Adapting a Technique.....	199
Table 71. Teachers' Developments about Assessment Technique.....	201
Table 72. Teachers' Developments in Purpose of Using Assessment Techniques	203
Table 73. Teachers' Developments in Writing Kinds of Test Items	204
Table 74. Teachers Developments in Content of a Test in Test Preparation	206
Table 75. Teachers' Developments in What to Consider while Writing Test Items.....	208
Table 76. Teachers' Additional Developments Regarding Test Preparation	210
Table 77. Teachers' Developments Considering Assessment Principles in Test Preparation.....	211
Table 78. Teachers' Developments in Use of Assessment Result.....	213
Table 79. Teachers' Developments in Sharing Assessment Results	215
Table 80. Teachers' Developments in Interpreting Assessment Results.....	217
Table 81. Teachers' Developments in Making Decisions Based on Assessment Results	219
Table 82. Teachers' Developments in Reflection of Assessment Result	220
Table 83. Teachers' Developments in Ethics	222
Table 84. Teachers' Developments in Administering Exams	224
Table 85. Teachers' Developments in Characteristics for Quality Assessment.....	226
Table 86. Teachers' Developments in What to Consider to Increase the Quality of Assessment	228
Table 87. Teachers' Developments in Relationships in the Issues Regarding Assessment ..	230
Table 88. Teachers' Developed Perceptions and Definitions of Assessment	232
Table 89. Teachers' Developments Regarding Assessment Type	234
Table 90. Teachers' Development of Assessment Purposes	235
Table 91. Teachers' Developments in Washback.....	237
Table 92. Teachers' Developments in Validity	239
Table 93. Teachers' Developments in Consistency.....	241
Table 94. Teachers' Developments in Grading in General	242

Table 95. Teachers' Developments in Reliability in Grading	244
Table 96. Teachers' Additional Developments in Grading and Scoring.....	246
Table 97. Teachers' Developments of Ethics in Grading.....	247
Table 98. Teachers' Developments in Ensuring Reliability in Administering Exams.....	249
Table 99. Teachers' Developments in Factors Influencing Reliability in Assessment	251
Table 100. Teachers' Developments in Knowledge of Measurement Error	252
Table 101. Teachers' Developments in How to Prevent Measurement Error.....	254



LIST OF FIGURES

Figure 1. Teacher cognition, schooling, professional education, and classroom practice	14
Figure 2. The steps taken in the analysis of the qualitative data	68
Figure 3. Emerging Themes Regarding Teachers' Nature of Cognitions About Language Assessment.....	84
Figure 4. Teachers' Strengths Related to Assessment	121
Figure 5. Teachers' weaknesses Related to the Assessment	151
Figure 6. Teachers' Developments Considering Language Assessment.....	191



LIST OF APPENDICES

Appendix A. Ethics Committtee Approval Of Bingöl University	300
Appendix B. Kwl Chart.....	301
Appendix C. Specific Dates For Overall Research Procedure.....	302
Appendix D. Reading List For Each Reflective Reading And Discussion Session.....	303
Appendix E. Reflection Report Template	306
Appendix F. Classroom Observation Form	307
Appendix G. Informed Consent Form	309
Appendix H. Permission For The Research.....	310
Appendix I. Semi-Structured Interview Questions For Exploration Of Teachers’ Current Nature Of Cognitions Regarding Language Assessment Literacy.....	311
Appendix J. Semi-Structured Interview Questions For Exploration Of Teachers’ Strengths And Weaknesses Regarding Language Assessment	312

CHAPTER I

INTRODUCTION

This chapter provides a brief overview of the background of the study. The problem statement, purpose, and research questions are then clarified. Furthermore, it offers the significance of the study and gives operational definitions employed in the study.

1.1. Background to Study

Teacher cognition research, exploring the unobserved aspect of teaching to show how teachers think and practice in the classroom, became prevalent in educational research in the 1960s. In comparison, teacher cognition studies in second/foreign language education began in the early 1990s (Tsui, 2011). Teacher cognition research has been based on the idea that teachers and teaching cannot be fully comprehended without knowing how teachers' thoughts, knowledge, and beliefs influence their actions (Borg, 2009). This called for "a more holistic and qualitative understanding of teachers' mental lives rather than an exclusive focus on observable behaviors" (Borg, 2006, p. 6).

Consequently, under teacher cognition research, researchers have investigated several facets of teaching in various circumstances, including pre-service and in-service teachers. These factors include instructors' decision-making strategies (Bailey, 1996), their cognitions and classroom practices regarding the teaching of grammar (Farrell & Patricia, 2005; Borg, 1999; Phipps & Phipps, 2009), and their metalinguistic awareness (Andrews, 2007), literacy instruction (Graden, 1996; Tercanlioğlu, 2001), and teaching writing (Burns, 1992; Ngo, 2018; Tsui, 1996).

Li (2017) mentions that learning to teach is a multifaceted process that includes making sense of teaching and growing as a professional in lesson planning, activity design, materials adaptation and development, instructional implementation of a plan, feedback, and assessment. According to Rea-Dickins (2004) assessment is an essential component of teaching because it necessitates evaluations regarding materials, learning tasks, and lesson content. Based on these explanations, the current study adopts the view that assessment is one of the teaching skills that teachers should develop to assist their instruction and students' progress.

The importance of assessment in teaching and learning cannot be questioned since teaching and assessment are viewed as two interconnected components in educational settings. As teaching and learning a foreign language has become a critical issue, language teachers have been increasingly engaged in language testing and assessment practices through which they assess their students more often than before during the language teaching processes (Fulcher, 2012). Not only for teachers but also for students good assessment procedures are essential since the quality of assessment is a requirement for the quality of instruction and learning (Stiggins, 1999). Therefore assessment is seen as an important component of education since it is viewed as a mechanism responsible for triggering learning (White, 2009) and provides feedback on the quality of instruction. In connection with this, assessment is beneficial to determine whether teaching is effective, whether students learn the desired behaviours, information, or skills, and the amount to which the objectives of a course are attained (Herrera & Macias, 2015; Rogier, 2014). As a result of the assessment, teachers, in turn, make judgments on content, materials, and alternatives to improve and implement successful teaching practices and instruction (Rea-Dickins, 2004). For this reason, teachers have come to the fore. Teachers are seen to be an essential element influencing teaching/learning. Their responsibilities, knowledge, and implementations extend beyond teaching. Understanding and practicing assessment are the teaching competencies that teachers should possess. As a result, teachers can be said to have two roles: teacher and assessor (Inbar-Lourie, 2013; Rea-Dickins, 2004; Wach, 2012). Language assessment is significant among these responsibilities because it governs teachers' decision-making processes concerning language teaching in accordance with language learning development (Inbar-Lourie, 2013; Oz & Atay, 2017).

It can be claimed that the main components of the assessment process are the teachers (Leung, 2014). Therefore, teachers who deal with assessment procedures on a daily basis are expected to keep up with the latest testing and assessment strategies and use them accordingly. Popham (2006) suggested that it is essential for teachers to have a certain level of assessment literacy. This is because the teachers are responsible for developing assessment processes, carrying out exams, scoring them, interpreting the findings, conveying assessment results and using them to inform educational decisions (Stiggins, 1999). A teacher with sufficient assessment knowledge and skills is also required to implement this knowledge into his/her assessment-related practices (Alkharusi, 2011). In short, a teacher cannot be effective at constructing tests, administering, scoring them, and interpreting the results if they lack

basic assessment knowledge. Considering these, foreign language instructors must be assessment literate to have these skills and carry out an efficacious assessment procedure.

Assessment literacy (AL) is a specific qualification of a teacher's professional competence. It is defined by Stiggins (1991) as having a fundamental knowledge of educational assessment and its associated skills. The importance of AL as a skill that all instructors must have is becoming more widely acknowledged (Popham, 2009; Xu & Brown, 2016). As part of assessment literacy, foreign language instructors must have a specific level of language assessment knowledge and abilities. Therefore, assessment training should be prioritized through in-service or pre-service education because 'teachers are not born testers' (Jin, 2010, p. 556).

However, to knowledge, in the light of the literature reviewed, studies focusing on teachers' assessment, especially language assessment knowledge levels, are generally descriptive studies that try to reveal the knowledge level of teachers within the scope of assessing four skills of either pre-service or in-service teachers (Ölmezer-Öztürk, and Aydın, 2018; Plake, 1993; Popham, 2009, Oğuz and Atay, 2017). Considering this, it might be said that there has been limited research investigating teachers' cognition about language assessment knowledge in language teaching especially focusing on its developments as a result of reflective reading and discussion. Therefore, this study intends to contribute to the literature on teacher cognition by exploring the perceived growth of English language teachers' cognitions about assessment especially language assessment knowledge and classroom practices via reflective reading and discussion.

1.2. Statement of the Problem

Previous research has usually examined the congruence between teacher and student beliefs (e.g., Cohen & Fass, 2001; Peacock, 1999), the influence of beliefs on teachers' instructional practices (e.g., Breen et al., 2001), and the changes in teachers' beliefs (e.g., Cabaroglu & Roberts, 2000; Mattheoudakis, 2007). Despite its broad reach (reviewed in Borg, 2003, 2015, and 2019), the study on language teacher cognition has devoted very little attention to the cognitive processes underpinning teachers' assessment practices (Yin, 2010). What teachers think, know, and believe may also significantly influence their lesson planning, activity and material design, evaluation and assessment of learning, and all other decisions made during the teaching process (Borg, 2003; 2015; Li, 2020).

Assessment and teaching are intertwined, as they mutually inform and influence one another (Malone, 2013). Language assessment literacy (LAL) helps instructors examine and evaluate student progress, uncover and analyse their assumptions, and promote learning (Scarino, 2013). Also, without LAL, teachers may be unable to assist their students in reaching higher levels of success which also results in a loss in educational quality (Howerton, 2016). In light of the importance of assessment to learning, there is an increasing demand for instructors to possess assessment knowledge, which will ultimately benefit students (Popham, 2011). On the other hand, several instructors expressed their concern that they do not have sufficient knowledge of assessment (Plake, 1993). Stiggins (2010, p.233) brought attention to this issue by asserting that "assessment illiteracy abounds." This claim suggests that teachers are charged with assessing their students, however it can be questioned whether or not they possess the necessary competence to achieve this. According to Popham (2009), instructors should evaluate students' competency and development, yet many lack fundamental knowledge of assessment terminology. As evidenced by the literature on assessment literacy, instructors lack enough assessment knowledge, and it is debatable whether they possess the skills necessary to evaluate students.

Since how teachers view the assessment and how well they implement it can influence how they teach and, consequently, how they assess (Jannati, 2015), language teachers as assessors should be examined first in terms of their language assessment knowledge because the entire assessment process is predicated on language teachers' expertise in language assessment. If language teachers are not skilled enough in language assessment, the entire assessment process will be influenced. In Türkiye, language teachers are expected to design, administer, score exams and analyse assessment procedures' outcomes. The issue is that they are expected to perform at a high level yet have little experience with assessment training. To go into detail, pre-service teachers only take a testing and assessment course and are not asked to assess students during practicum. Additionally, language instructors are not required to attend assessment-related professional development programs. The situation becomes more complex when language instructors working at higher institutions who are not graduates of the English Language Teaching department assess the students. These language teachers do not even take any course concerning language testing and assessment during their pre-service training. Furthermore, they do not need to possess a formation education license to be recruited as a lecturer at any English preparatory programs at universities. With this inadequate training in assessment, graduates begin working as English language instructors.

Teachers with insufficient assessment knowledge and background experience also undertake the assessment and evaluation burden if there is no testing unit in the foreign language schools in higher education in which they work. The question then arises regarding how informed or competent instructors are in assessing students. As a result, there is a dearth of research in language assessment literacy to shed light on this issue. In Türkiye, there are few studies on determining EFL teachers' language assessment knowledge working at a foreign language school (Ölmezer-Öztürk, and Aydın, 2018; Oğuz and Atay, 2017; Büyükkarcı, 2016; Mede and Atay, 2017; Hatipoğlu, 2015). Such kind of investigations are crucial because teachers' needs may be explored by identifying their strengths and weaknesses in language assessment. Based on their needs, testing and assessment courses in pre-service education and teacher professional development programs for in-service teachers can be designed, established, and delivered. Therefore, it might be pointed out that there is a need in Türkiye to investigate the EFL instructors' language assessment knowledge as well as their needs considering the issue. As a starting point, how teachers think, know, believe and do considering assessment should be determined. Additionally, their language assessment strengths and weaknesses might be clarified.

1.3. Significance of the Study

As teachers' assessment knowledge affects the quality of instruction, they should have essential assessment skills and knowledge to meet their teaching successfully and students' learning needs, thus contributing to the growth of both students' and teachers' performances (Howerton, 2016; Huang & He, 2016; Khadijeh & Amir, 2015). Based on this, extensive research on teachers' language assessment knowledge might be needed. As a result, findings might improve the field of such research and provide novel ideas that shed light on issues about improving quality assessment and teaching. To our knowledge, this study may pioneer to similar studies to be conducted in this field. It began by investigating teachers' current perceptions of LAL. The strengths and needs, namely their weaknesses, of teachers were then identified, and a series of RRD sessions were designed and implemented. Finally, the development of teachers' cognitions of LAL was discovered. Considering this, the current study attempted to analyse a thorough investigation, understanding and evaluation of teachers' cognitions of LAL since it not only assessed teachers' knowledge but also their LAL practices were examined.

The current study hopes to contribute to language assessment research in many ways. First, exploring what EFL instructors working at preparatory programs think, know, believe and do about language assessment would be helpful in clearly identifying their nature of cognitions of assessment which in turn lead to their weaknesses and strengths to be discovered. Following that, findings related to language teachers' weaknesses and needs regarding language assessment may provide insights that can be used for adaptation and improvement in assessment courses provided in pre-service teacher education programs, believing Deluca and Klinger's (2010)'s idea that pre-service education is important for the future professional life of teachers. In addition, in accordance with Borg's (2003, p. 106) request for "making actual data from the body of research on teacher cognition and practices available to trainees and teachers as the basis for teacher education activities," we can suggest that the findings may be beneficial in identifying ways to enhance pre-service assessment training (Malone, 2008). That is, this study is thought to offer some insights into how to improve this literacy and how to design and arrange language assessment training at universities, which will assist pre-service teachers in constructing their LAL. Last but not least, the key contribution of the study may be its identification of the relationship between language assessment training and the development of LAL through in-service teacher reflective practices. In light of the findings, in-service professional teacher education programs for teachers can be designed, and teachers may benefit from them in terms of language assessment. Furthermore, the results of language teacher cognition (LTC) research can significantly advance teacher education. The findings about professional development can help shape institutional policies in that administrators may plan to hold appropriate in-service teacher education programs.

In addition, recent research shows the value of concurrently researching teacher cognition and teacher learning. These studies show the impact of prior learning experiences, previous coursework, educational background, and sociocultural settings on pre-service teacher cognition and how teacher education programs may shape student teachers' views, attitudes, and knowledge via reflection (Johnson, 2018; Kubanyiova, 2015). Thus, the present study aims to offer significant implications for teacher learning by examining teacher cognition; consequently, it may provide a cognitive understanding of teacher learning (Freeman, 1992, 1993).

Studies researching the changes in teacher cognition on a longitudinal basis (Cabaroglu and Roberts, 2000; Özmen, 2012) are limited in number. A short-term research may not be

enough to explore the development and change in teacher cognition when working with a group of in-service teachers in an English as a foreign language context. According to Sert (2015), a longitudinal research design is required to track the changes in teacher cognition. As Borg (2015) stated, longitudinal investigations on the change in teacher cognition would contribute to the existing literature a lot as the research agenda is limited. Furthermore, Birello (2012) indicated that longitudinal research designs are important in understanding teacher learning over a period of time. In line with these suggestions, it can be claimed that there is a need for further research investigating the processes through which changes in language teachers' cognitions and practices occur. Accordingly, by investigating in-service teachers' construction of language assessment literacy cognition with a longitudinal qualitative research design, the present study will try to address the gap in the literature.

The final value of this study is related to the limited research focusing on the development of EFL instructors' cognitions regarding language assessment knowledge in a preparatory program context in Türkiye. Numerous investigations regarding EFL/ESL teachers' cognitions have been carried out worldwide to explore teachers' beliefs and ways of thinking about language testing and assessment (Mede and Atay, 2017; Hatipoğlu, 2015; Seferoğlu, Korkmazgil, & Ölçü, 2009). However, such studies are quite limited in the Turkish context. Most of the studies in Türkiye are conducted in pre-service years with student teachers. Most of the time, those studies are being carried out to shed light on pre-service or in-service teachers' knowledge, beliefs and practices about assessment (Mede & Atay, 2017; Hatipoğlu, 2015; Büyükkarcı, 2016; Ölmezer-Öztürk and Aydın, 2018). The studies on in-service teachers' cognitions, especially about cognition development, are limited. Considering these limitations, this study aims to reach a broader picture and fill in this gap in the literature.

1.4. Purpose of the Study

This current study firstly aimed at identifying EFL instructors' nature of cognitions regarding language assessment. Also, it aimed at exploring teachers' weaknesses and strengths considering language assessment. Based on the results of such procedures, it aimed to plan, design and implement a set of RRD sessions in which teachers share, reflect and help one another's learning. Finally, development in teachers' cognitions considering language assessment was tried to be found out. Based on these purposes, the following research questions were formulated for the current study:

1. What is the nature of EFL instructors' cognitions regarding language assessment?

2. What are participants' strengths and weaknesses considering language assessment, as they perceive?
3. Do reflective reading and discussion sessions lead to any development in EFL teachers' cognitions concerned with language assessment, as they perceive? If so, what are the changes?

1.5. Operational Definitions

Assessment: "A broad term meaning a process for obtaining information that is used for making decisions about students; curricula, programs, schools; and educational policy" (Brookhart & Nitko, 2008, p. 4).

Language Assessment: It is gathering information and deciding how well a language learner knows and can use a language (Chapelle & Brindley, 2010).

Language Assessment Literacy: Having assessment-related knowledge, abilities, and principles (O'Loughlin, 2013; Lam, 2015).

Teacher Cognition: It is defined as "the unobservable cognitive dimension of teaching – what teachers know, believe, and think" (Borg, 2003, p. 81).

Case Study: Case studies "investigate a contemporary phenomenon within its real context when the boundaries between phenomenon and context are not clearly evident and in which multiple sources of evidence are used" (Yin, 1994, p. 23).

The Zone of Proximal Development (ZPD): "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).

The Zone of Teacher Proximal Development (ZTPD): "The distance between what teaching candidates can do on their own without assistance and a proximal level they might attain through strategically mediated assistance from more capable others" (Warford, 2010, p. 253).

Reflection: Reflection is the practice of professionals learning from their experience and becoming aware of their implicit knowledge base. (Schön, D. , 1987).

CHAPTER II

LITERATURE REVIEW

This chapter will provide an overview of the research literature related to teachers' cognition of assessment especially language assessment and its development as a result of reflective reading and discussion. First, the historical development of teacher cognition, its nature, the factors affecting it and its relationship with teacher practice will be discussed. In other words, with examples from the current literature, the evolution of teacher cognition and the mutual interaction among teacher cognition, classroom practices, and teacher learning will be discussed. The current study also will present studies examining if and to what degree language instructors' cognitions and behaviours correspond to one another. Secondly, the notions of assessment, assessment literacy and language assessment literacy will be explained followed by a basic overview of the research on these topics. Finally, the study's theoretical framework, which is Vygotskian sociocultural theory, will be explicated.

2.1. Language Teacher Cognition

In the 1970s, behaviourist techniques dominated the field of general educational research. In the 'process-product' research, the empirical focus was placed on examining successful teaching behaviours (Freeman, 2002). The objective was to identify and share good teaching behaviours so that they might be included in the training, assuming that successful learning would result. The frequency of occurrences of specified categories of behaviour in classroom observations served as the basis for research. However, this behaviourist theoretical viewpoint was criticised for being reductionist, category-restrictive, and unidirectional (Erickson, 1986). This methodological discontent prompted researchers to develop theories for teacher cognition that influence their behaviour. Then studies on teachers' thought processes began to appear, and most disciplines revealed a transition from a product-oriented (behavioural) paradigm to a process-oriented (thinking) paradigm. These improvements in teaching made it possible to start studying how teachers think. As a result, researchers began using the term teachers' mental lives, which was coined to increase understanding of classroom learning by focusing on the "mediational linkages" (Walberg, 1972, p. 33) that form the core teaching processes and learning. In this way, the emphasis of studies in education shifted from teaching efficiency, students' classroom behaviour, and student accomplishment to teachers' mental life, with a concern for the influence of thinking on

behaviour. LTC scholars remained research-intensive while realising the need for better conceptual and methodological clarity and the paradox of examining the unobservable realm of teachers' mental lives. Interest was generated by expecting discoveries to lead to improved practice, more informed teaching and learning procedures, and clarified theory.

Research activities have resulted in a proliferation of terminology in the literature (Pajares, 1992). Many researchers looked at teacher cognition from different points of view, and many terms have been created in the field. Definitional differences in the researchers' terminology complicate the procedure, even though they all study the same construct. Researchers with an epistemological perspective have developed different explanations regarding teacher cognition. For example, Elbaz (1981) put forward the term "practical knowledge" since she argues that teachers' knowledge is mostly derived from their teaching experiences. According to her, there are situational, theoretical, personal, social, and experiential orientations within the framework of practical knowledge. To illustrate more, teachers relate their knowledge to a particular situation and its tasks through situational orientation. As for theoretical orientation, teachers may reject or embrace a specific theory or develop theories of practice. Therefore, regardless of the teacher's position, the way that teacher approaches ideas will govern the kind of theoretical knowledge that the teacher will employ and how. Also, teachers represent their aim and personal meaning via personal orientation in their interactions with students. Social orientation demonstrates how teachers utilise their knowledge and expertise to organise their social circumstances. Finally, Elbaz (1981) defines experiential knowledge as "the teacher's knowledge grows out of the world of teaching as she experiences it; it gives shape to that world and allows him to function in it" (p. 58). Another researcher tried to explain teachers' knowledge using different terminology. Golombek (1998) coined a new term, teachers' personal practical knowledge. According to Golombek (1998), personal practical knowledge serves as an interpretive framework that guides teachers' activities and helps them make sense of their classrooms. Another researcher, Gatbonton (1999), proposes the idea of teachers' "pedagogical knowledge" by explaining it with six broad areas. They are "knowledge of handling language items, factoring in student contributions, determining the contents of teaching, facilitating the instructional flow, building rapport and monitoring student progress (Gatbonton, 1999, pp.42-44). Additional examples from terms which were employed in language teacher cognition research are practical knowledge (Gholami & Husu, 2010; Meijer, Verloop, & Beijard, 1999), theories for practice (Burns, 1996), knowledge about language (Bartels, 2009; Borg, 2005),

epistemological beliefs (Flores, 2001; England, 2017), and apprenticeship of observation (Moodie, 2016).

In conclusion, the researchers who emphasise the concept of "knowledge" in their studies do not only refer to knowledge acquired through professional training and development; they also emphasise the personalised nature of teachers' knowledge, which is shaped by teaching experience in multiple domains within contextual conditions.

The construct of teacher cognition is so closely interwoven with diverse concepts and viewpoints that Borg (2003), in his fundamental overview of teacher cognition research, notes that this terminological variety "should not mask the considerable overlap which exists among them" (p. 83). Likewise, Woods and Çakır (2011) note that although there may be a distinction in the sense that they are accounts of separate theories, they may otherwise be the same. In contrast, the fact that two scholars use the same phrase does not imply that they share the same ideas.

This section explains the emergence of language teacher cognition, terminology confusion related to it, and terms coined by researchers referring to language teacher cognition. The following sections will discuss dimensions of language teacher cognition, factors that shape language teacher cognition, and studies based on LTC.

2.2. Dimensions of Language Teacher Cognition

This section will describe the dimensions of teacher cognition and how they influence the classroom practices that teachers select and perform. These dimensions include; teacher understandings, teacher beliefs and teacher experiences. Each of them will be explained subsequently.

2.2.1 Teacher Understanding/Knowledge

Teacher knowledge will be clarified as the first dimension. Teacher knowledge is regarded as the key component in research based on teacher cognition. This is because the decisions teachers make in class will depend on how much they know about a particular topic or how little they know about it. So, the question is, what do the teachers know? In this current study, "teacher understanding" or "teacher knowledge" refers to what teachers know about assessment, and the terms will be used interchangeably. Grossman and Richert (1988) define teacher knowledge as "a body of professional knowledge that encompasses both knowledges of general pedagogical principles and skills and knowledge of the subject matter

to be taught” (p.54). This teacher’s knowledge is dual. It consists of “practical knowledge” which entails an understanding of subject matter, instructional routines, classroom management, and student requirements, and 'personal aspect,' which is the instructors' gained knowledge of themselves via their experiences (Kiss &Lin, 2016). It is the knowledge which is combined with expertise that leads to learning. Therefore, it can be said that the role and contribution of expertise in teachers' learning cannot be denied. In other words, the knowledge gained by teachers and their experiences affects their understanding of any subject and presenting it to students in the classroom.

2.2.2 Teacher Beliefs

The second component of a teacher's cognition is their set of beliefs. Beliefs help us to understand how teachers conceptualise their work. Then the question that is “what are beliefs?” emerges. Beliefs have been defined by various researchers based on their perspectives. Kagan (1992) indicates that "teacher belief' is not used consistently, with some researchers referring instead to teachers' "principles of practice," "personal epistemologies," "perspectives," "practical knowledge," or "orientations." (p.66). Also, she defines beliefs as unconscious beliefs and expectations regarding the students, classrooms, and the subject matter that will be covered in teaching (Kagan, 1992). Beliefs are also a collection of strong emotions and attitudes instructors hold towards matters that might influence the teaching-learning relationship.

The way teachers behave in class will be affected by these beliefs. Studies have demonstrated that teachers' beliefs influence how they behave in the classroom and are indicators of what a teacher gives (Morina, 2016). Studies also reveal that teachers' beliefs might be difficult to alter since they are formed early in life (Borg, 2003). Therefore, professional development programs or interventions to develop teachers' cognitions may be recommended to comprehend and account for teachers' beliefs. Considering this, in this current study, the perceptions of participating teachers about assessment will be explored first.

2.2.3 Teacher Experiences

The third dimension of teacher cognition is teacher experiences. Experience influences the kind and content of teachers' decisions during lessons (Ozturk, 2015). These experiences are made up of teachers’ previous language learning experiences, their pre-service teacher education experiences, their experiences in the years spent as an in-service language teacher, and their teaching practices and these experiences influence the formation of their cognition.

The first component of teachers' experiences is their learning experiences while learning the language. According to Lortie (1975), when students start their university education, they already know how language is taught. In other words, before enrolling in college, students spend thousands of hours observing and assessing the actions of their teachers. This is called "apprenticeship of observation" (Lortie, 1975). The apprenticeship of observation considers a teacher's personal life experiences as a student. Borg (2015) also claims that teachers' earlier language learning experiences establish cognitions that serve as the foundation for their early conceptions of language instruction throughout teacher education and may continue to impact them throughout their professional life. However, Boyd et al. (2013) claim that teacher education programs challenge the apprenticeship of observation by recognising students' personal experiences and finding a means to mediate them in some way through coursework. Student teachers' preconceived notions about teaching languages can either be solidified or completely dismissed at this stage. Although it is asserted that beliefs are resistant to change, it is also true that they alter due to the incorporation of prior thinking and ideas (Morina, 2016). One way of manipulating teachers' implicit beliefs, which are claimed to remain unchanged, is the education they receive during their university years. These programs must be designed to promote the professional development of pre-service teachers and enhance their practices in foreign language classrooms. The final area of the effect is the actual setting in which teachers teach the language. Teachers' classroom practices are affected by various interactions and often compete for institutional, pedagogical, personal, and physical elements (Borg, 2003). It can be said that these indirectly affect teachers' experiences in the classroom and, subsequently, their language teacher cognition formations.

2.3. Factors that Shape Language Teacher Cognition

Since the emergence of the study on teacher cognition, educational researchers have focused their attention on various factors that affect teachers' beliefs, thought processes, and knowledge. Borg (2003) summarises the data of the research he examined and concludes that "a wide range of interacting and often conflicting factors shape language teachers' cognitions and instructional practices" (p. 91). Borg (2006) developed a framework that depicts how teacher cognition evolves and the interaction link among teacher cognition, classroom practices, and teacher learning based on the notion that the area of language teacher cognition lacks a unified and systematic research agenda. According to the model, teachers' theories, beliefs, and knowledge regarding teaching, teachers, and learning constitute their cognitions. Moreover, cognition about teaching, teachers, learning, students, subject matter, curriculum,

resources, instructional activities, and self consists of particular constructions such as beliefs, knowledge, theories, assumptions, images, metaphors, conceptions, and perspectives. In this sense, the phrase "teacher cognition" can be utilised as an umbrella word for these numerous structures. When the framework is investigated, it can be said there is such a two-way reciprocal interaction between teacher cognition, professional coursework, and teacher cognition and classroom practice that they mutually impact one another. Conversely, schooling affects teacher cognition and professional coursework with a filtering effect on these constructs. Early cognitions from prior schooling shape recently introduced information in the teacher education program (Bruner, 1996). Finally, it is demonstrated that contextual elements affect teacher cognition and classroom practice on the model, as the characteristics of the context influence both the act of teaching and thinking about it.

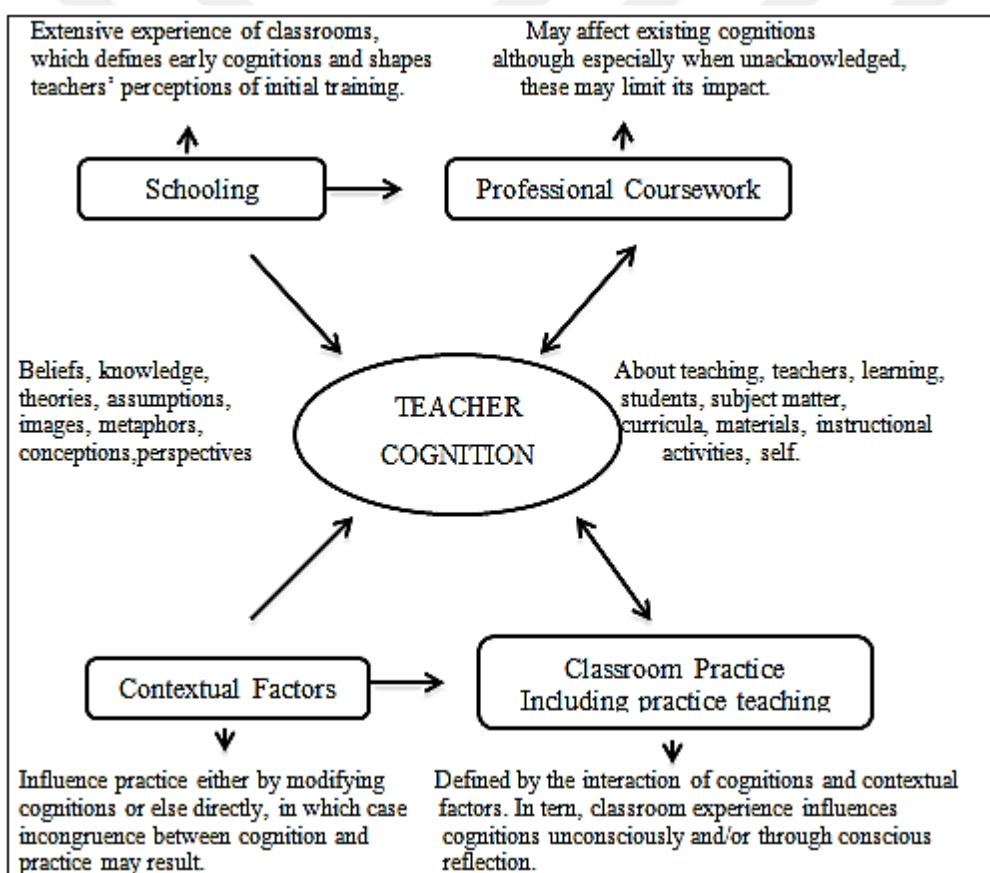


Figure 1. Teacher cognition, schooling, professional education, and classroom practice (Borg, 1997, cited in Borg, 2003).

This schematic model represents how instructors' experiences as learners influence and shape their cognitions. It can be said that practices that teachers apply in the classroom are influenced by cognition. Still, cognition is affected by the circumstances in the classroom setting. In this model, teacher cognition and contextual variables interact to determine

classroom practice. Borg (2003) indicated that contextual factors moderate teacher cognitions and practices, allowing teachers to conduct teaching consistent with their cognitions.

In accordance with Borg's model, Urmston (2003) conducted longitudinal research and discovered that pre-service English language teachers' experiences as students and the time they spent in classrooms for practice teaching significantly impact their beliefs and knowledge about teaching English. Farrel (2008) identifies learning to teach as a complicated procedure and provides a list of variables that affect first-year teachers. He admits the effect of prior schooling and teacher education programs and the importance of first-year socialisation inside an existing school culture.

2.4. Relationship between Language Teacher Cognition and Classroom Practices

It appears difficult to figure out that “language pedagogies are based on teachers’ accounts of how they work without reflecting upon *actual instances* of practice” (Breen *et al.*, 2001, p. 498). However, there is not always a direct relationship between instructors' knowledge and actions. Due to potential causes and limitations, it is unreasonable to expect instructors' views and practices to align constantly with one another. Teachers may not put their views into practice as these variables are mostly beyond their control. Considering this, Basturkmen *et al.* (2004) put forward that it could make more sense to look at what extent teachers’ views are represented in their practices rather than trying to ascertain whether or not their practices reflect their beliefs.

It has been revealed that teachers' classroom practices are influenced by various interacting and often competing elements (Borg, 2003). Likewise, it is stated that teachers' knowledge and views substantially impact how they plan lessons, deliver instructions, and interact with students (Richards & Lockhart, 1994).

Through teachers' checking out what they do in the classroom and its subsequent reflection on their cognitions, the link between cognitions and practices is perceived as interactive and affecting one another (Breen *et al.*, 2001). Clark and Peterson (1984) remarked that teacher cognition and behaviour are mutually related by adding that “ teachers’ actions are in large part caused by teachers’ thought processes, which in turn affect teachers’ actions” (p. 13). Similarly, Foss and Kleinsasser (1996) assert that there is a "symbiotic relationship" (p. 441) between the thought processes of pre-service teachers and the way they teach. Borg (2006), on the other hand, views the connection between cognition and practice as "neither linear nor unidirectional." His additional explanation is as follows:

It is not linear because cognitions and practices may not always concur due to the mediating influence of contextual factors. It is not unidirectional because teachers' cognitions themselves are shaped in response to what happens in the classroom. Language teaching can be seen as a process defined by dynamic interactions among cognition, context and experience (p. 275).

Borg (2006) expresses his concern that a study focusing on the relationship between teachers' cognition and their practices that is conducted in isolation from the environments in which teachers work—specifically social, institutional, instructional, and physical settings—may result in overly simplistic conclusions. For this reason, participants were observed in their natural teaching environments in this current research. The study aims were tried to be achieved without separating instructors from their teaching context.

As stated earlier, teacher cognition includes beliefs and knowledge. Studies try to reveal whether teachers' beliefs match their classroom practices. While some show a match between teachers' beliefs and actual classroom practices (Öztürk, 2014; Johnson, 1992), some studies have found that teachers' beliefs and in-class practices mismatch for various reasons (Peaeson, 1985; Philips & Borg, 2009). Li (2013) characterises the link between ideas and practice as very complicated, given that teachers' views encompass several aspects, such as students, curriculum, learning and instruction. These concepts are linked with classroom practice and cannot be studied individually (Li, 2008). The studies conducted focusing on language teacher cognition in the literature will be mentioned in the following section.

2.5. Research in Teacher Cognition

Research on teacher cognition, which includes what teachers think, know, and believe and its relation to teachers' educational practices, has drawn more attention in recent years. As a result, several studies have been conducted to learn more about teachers' knowledge, beliefs, thoughts and practices. However, greater focus was placed on the 'belief' component, resulting in a vast quantity of educational research about the study of teacher belief. Much research underlined the relationship between teacher beliefs and educational practices (Breen, 1991; Burns, 1996; Calderhead, 1996; Johnson, 1992, 1994; Pajares, 1992; Richards & Lockhart, 1996; Smith, 1996; Thompson, 1992). Literature provides numerous indications that teachers' beliefs may: (a) be strongly influenced by teachers' own experiences as learners (Holt Reynolds, 1992); (b) outweigh the influence of teacher education programs on what teachers perform in the classroom (Kagan, 1992; Richardson, 1996); (d) not always be

reflected on what teachers practice in the classroom (Dobson & Dobson, 1983; Pearson, 1985; Tabachnick & Zeichner, 1986); and (e) interact mutually with their practices (Richardson, 1996). The study of teacher beliefs has become more relevant in educational research on teacher cognition. However, research on teacher cognition cannot be limited to the study of teacher belief because Borg (2003, p. 81) defined teacher cognition as "unobservable cognitive dimensions of teaching - what teachers know, believe, and think," and it has been associated with complex (Borg, 2005; Gabillon, 2012), changeable (Flores, 2001; Gabillon, 2012; Johnson & Goettsch, 2000), and influential (Burns, 1992; Johnson, 1992) characteristics.

2.6. Research on Language Teacher Cognition

Early in the 1990s, research on the education of second language teachers began to emerge, and it has grown significantly since then (Borg, 2009). This section presented a literature review on the cognitions of English teachers in the pre-service years when they received teacher education, their cognitions in the initial years of their career, and their cognitions in in-service years.

Substantial research on language teaching has focused on teacher cognition in the setting of pre-service teacher education. Several studies sought to understand and depict the cognitions of pre-service language teachers, while others focused on determining the impact of teacher education programs on student teachers' cognitions. In addition to these, several studies highlight the significance of student teachers' "prior beliefs" that are brought by them to their pre-service teacher education programs. First, Wray (1993) examined the knowledge and beliefs of prospective language teachers in the United Kingdom and found that teacher candidates' comprehension of grammar was insufficient. Next, through narratives, Johnson (1994) sought to investigate pre-service teachers' ideas about language teaching and language instructors to determine how these beliefs affected teachers' instructional performances. Johnson (1996) investigated pre-service teachers' views on their initial teaching experiences and discovered conflicts and uncertainty. Numrich (1996) also investigated prospective teachers' diaries throughout their practicum. Common themes concerning language teaching and learning and their opinions of their personal needs emerged, and related findings were presented accordingly. Warford and Reeves (2003) examined pre-service teachers' metaphors regarding language teaching and reported their views on the issue.

Also, Maloch et al. (2003) conducted a study with pre-service EFL teachers on reading instruction and described participants' beliefs on the best way to teach reading. In addition to these studies that have focused on describing the nature of prospective teachers' cognitions on a specific topic, others have sought to examine how their cognitions have evolved during and as a result of pre-service teacher education. It was generally believed that pre-service teachers in teacher education programs were in the process of constructing their pedagogical beliefs and practices (Ertmer & Ottenbreit-Leftwich, 2010). As they learn and are equipped with professional and pedagogical knowledge, pre-service teachers adjust their prior beliefs about teaching and construct how their teachings will be (Hall, 2005). Cabaroğlu and Roberts (2000), Grisham (2000), Johnson (1994), Peacock (2001), and Yook (2010) investigated pre-service teachers' cognition development during or as a result of their teacher education programs to see how teacher education play a role in constructing or changing prospective teachers' cognitions. While some of these studies from the literature show that pre-service teacher cognitions are limited or not affected by teacher education programs, some of them revealed that teacher education programs have significant effects on pre-service teachers' cognitions. For instance, Peacock (2001) investigated the second language learning beliefs of pre-service EFL teachers and discovered that the courses taken by student teachers in their teacher education programs did not affect their beliefs, namely their cognitions. Cabaroğlu and Roberts (2000), Sendan and Roberts (1998), and Tüzel and Akcan (2009), on the other hand, highlighted the strong impact of training programs or courses on student teachers' knowledge and beliefs. In addition, several research focused on the cognitions student teachers brought from earlier learning experiences to teacher education programs and the influence of such cognitions on their developing language teaching understandings and practices. Student teachers enter the classroom with their pre-existing cognitions about teaching and learning, and they continuously reconstruct their beliefs about language learning and teaching as a result of personal experiences (Loughran & Russel, 1997). However, sometimes, those cognitions remain stagnant to change or may include inappropriate and unrealistic perceptions of teaching and learning (Tatto, 1998; Brookhart & Freeman, 1992). Considering this, Borg (2006) states that student teachers' previous language learning experiences help them develop their initial cognitions about learning and language learning while entering their teacher education years, forming the basis of their beliefs about second language teaching. Similar to the studies stated above, research has been conducted on pre-service teachers' pre-existing knowledge about teacher education. Numrich (1996) examined how past experiences related to pre-service teachers' classroom practices and discovered that

student teachers avoid performing specific instructional practices owing to unpleasant experiences they had as learners. Also, Williams and Burden (1997) highlighted the significance of prior experiences and deeply formatted beliefs about language learning in the construction of beliefs, arguing that they may have a greater impact on pre-service teachers' classroom performance than a specific topic that they learned in the teacher education program. It could be concluded that, as Decker and Rimm-Kaufman (2008) elaborated, student teachers enter teacher education programs with a set of beliefs about teaching that stem from previous educational experiences and that these initial beliefs about teaching significantly shape their subsequent beliefs about teaching.

In addition, studies on teacher cognition have also focused on novice teachers. These studies generally revolve around describing the novice teachers' nature of cognition and comparing novice-experienced teacher cognitions. Abdullah-Sani (2000) explored the beliefs about teaching and learning held by Malaysian novice teachers in a longitudinal research conducted firstly when the participants were student teachers and at the beginning of their teaching careers. According to emerging results, previous learning and education let pre-service teachers create and develop beliefs that governed them during their first years of teaching. As for the studies comparing the cognitions and practices of novice and experienced teachers, Cumming (1990) investigated the decision-making procedures of novice and experienced language teachers and discovered statistically significant differences in how they developed criteria for organisation and content, addressed language errors, and assessed the quality of the works while they were evaluating written compositions. In a similar vein, Tsui (2003) identified several distinctions between experienced and inexperienced teachers in a study based in Hong Kong. An experienced language teacher, for instance, has extensive and integrated knowledge of the language, language teaching, language learning, and students' interests. In contrast to experienced teachers, novice teachers were found to lack some educational practices, such as handling unexpected classroom situations. Although novice and experienced teachers appear to have similar beliefs, it has been discovered that their classroom performances differ (Seferoğlu, Korkmazgil, & Ölçü, 2009).

Studies on in-service English teachers' cognition have focused on a variety of topics, including teaching specific language regions and teaching language skills. Burns (1996) investigated the beliefs and practices of English language teachers in an Australian classroom of adult language learners. The emerging findings suggested that organisational requirements and institutional context influenced how the teachers plan their lessons and determine the

content accordingly. Smith (1996) examined the connection between teachers' beliefs, instructional decisions, and contextual factors in another investigation of English language teachers in Canada. This study distinguished between planned and unplanned interactive decisions, asserting that student or teacher-related variables prompted unanticipated decisions. Some research has focused on the language teachers' views and beliefs regarding integrating technology in language classrooms (Lam, 2000) and computer-mediated language instruction (Lawrence, 2001). Flores (2001) undertook survey research to investigate the views of 176 instructors and discovered that the beliefs were never static. The participants' beliefs reflected the significance of language and culture in teaching. Also, studies on teachers' cognitions considering certain language teaching approaches were carried out. For instance, Sato and Kleinsasser (1999) used a survey, interviews, and observations to investigate the practical understanding of ten language teachers about communicative language teaching (CLT). Despite the participants' strong attitudes toward communicative language teaching, there was no indication of it in the actual teaching analysis. Choi (2000) offered additional support by focusing on in-service EFL teachers' views of CLT in Korea and identified differences between their beliefs and instructional practices. Aside from teacher cognition studies on the match/mismatch between instructors' cognitions and their actual classroom practices, many teacher cognition studies were also on how to teach grammar. These studies primarily sought to investigate instructors' perceptions of the language and the use of explicit grammar in foreign language instruction. In another study, Richards, Gallo, and Renandya (2001) used a self-reported questionnaire to explore the self-expressed views and actual classroom practices of in-service EFL teachers on grammar teaching and figured out a mismatch between teachers' stated beliefs and their actual classroom practices. Many teachers displayed communicative techniques in their instruction, despite their beliefs emphasising the significance of explicit deductive grammatical teaching. Mangubhai et al. (2004) examined the practical theory of a teacher about communicative language teaching as part of a case study. As a result, the teacher was found to establish a well-formed practical theory, which was also reflected in her classroom practices. Hislam and Cajkler (2005) conducted a case study to observe how in-service teachers used teacher knowledge of grammar in their practice. They found that the use of language knowledge was challenging and that the participants' use of traditional sources like books and websites was insufficient when teaching grammar. This study also demonstrated that inexperienced teachers struggle to learn and internalise grammatical terms.

Gatbonton (2008) conducted a study with novice and experienced teachers in Canada to investigate their thinking when teaching English as a Second Language (ESL). It was found that the pedagogical knowledge and practices of experienced teachers were more stable since they had more experience in teaching. In his longitudinal research, Borg (2011) sought to determine how an intense eight-week in-service language teacher practising program influenced practising teachers' ideas about language teaching and learning in the United Kingdom. As a result, there was a significant influence on the participants' emphasis on strategies for developing classroom practices and changes in their previous beliefs. In a case study with two non-native English-speaking student teachers in the United Kingdom, Li (2012) investigated how these Chinese instructors create and develop beliefs about subject content, learning, teaching, learners, and the teacher throughout a one-year training program. It was proposed that teacher education programs influenced and developed pre-service teachers' ideas, resulting in identity changes among student teachers. Tsang (2004) investigated the importance of personal practical knowledge in the interactive decision-making processes of prospective teachers in Hong Kong, including previously established ideas and theories and those generated throughout teacher training programs. Nishino (2008) conducted an exploratory study to explore the beliefs and practices of Japanese EFL instructors regarding communicative language teaching. The teachers were found to have excellent knowledge of communicative language instruction as well as a good understanding of the roles of learners and teachers in a communicative classroom. However, several issues were claimed to have prevented the implementation of communicative language teaching. Participants stated that improving the classroom environment requires effective communicative language teaching practices. Goker (2006) wanted to know if pre-service EFL teachers would exhibit a development in their teaching and self-efficacy after undergoing a peer coaching training program. He discovered significant differences between pre-service teachers who had received traditional training and those who had undergone a peer coaching program in terms of setting objectives, restating points, using examples, repeating items, asking questions, student questions, and practising time. Mattheoudakis (2007) conducted longitudinal research to explore pre-service EFL teachers' views about learning and teaching and any changes in those beliefs resulting from a teacher education program. The findings revealed a considerable progressive shift in student instructors' opinions throughout the curriculum. The results were evaluated in light of the structure and setting of the particular teacher preparation program.

2.7. Studies on EFL Teachers' Cognitions in the Turkish Context

Studies on teachers' cognition from various topic areas in the Turkish setting are presented in this section. To begin, Sendan and Roberts (1998) reported on the case of a single student teacher's personal theories about quality instruction and how those theories evolved. According to the study, changes in the participant's thinking were complicated rather than linear, as the structure of his personal theories was deconstructed and rebuilt over a 15-month period. Next, Tercanlioğlu (2001) investigated pre-service EFL teachers' perceptions of themselves as readers and potential reading instructors and discovered evidence that the participants in the research were not passionate about reading and were unsure of their effective reading abilities. They did, however, feel that excellent reading teachers should also be good readers. Üstünel (2008) explored the link between novice instructors' classroom management beliefs and behaviours. The data gathered from questionnaires, discussion sessions, and classroom observations revealed that the more teaching experience a novice teacher had, the better his/her views were reflected in classroom practices because participants found it difficult to put their views into practice when dealing with classroom discipline in the first semester. In a Turkish pre-service education program, Balçıkanlı (2010) examined student teachers' perceptions of learner autonomy; the students exhibited good attitudes toward the learner autonomy principles. However, most prospective teachers did not want students to participate in decisions on the time and location of classes and the texts to be utilised. Özmen (2012) conducted a four-year longitudinal study on the beliefs of pre-service EFL teachers about language learning and teaching during their teacher education program and analysed the program's influence on the changes in beliefs. It was observed that changes of varying degrees appeared at different phases of the program. Specifically, pre-service teachers' practicum experiences had a greater influence on the formation of beliefs regarding language learning and teaching. Seferoğlu, Korkmazgil, and Ölçü (2009) investigated pre-service and in-service teachers' cognition about teachers using the metaphor elicitation approach. The data was gathered from three different groups of language teachers: junior pre-service teachers, senior pre-service teachers, and in-service teachers. The findings revealed that nearly all groups of participants perceived the 'teacher' as a 'guide.' The percentage of in-service teachers who saw themselves as "facilitators" was greater than the percentage of pre-service teachers. It was argued that when teachers gain experience, they adopt a learner-centred approach. Cabaroğlu and Yurdaisik (2008) explored university preparatory school teachers' perspectives on teaching reading and using reading strategies in class in three

different institutions. According to the findings, dealing with unknown terminology and themes was one of the most challenging areas of reading instruction. On the other hand, teachers employed more pre-reading techniques than post-reading strategies, and participants who utilised reading strategies in their daily lives used them more in class.

The following parts of this chapter respectively provide detailed information regarding key terminology in testing and assessment, types of assessment, principles of language assessment, assessment literacy and language assessment literacy. Afterwards, the existing studies conducted on assessment literacy and language assessment literacy are presented.

2.8. Assessment in Language Teaching

Assessment and teaching are two components of education that interact with one another. They cannot be separated since "assessment and teaching are two sides of the same coin" (DiRanna et al., 2008, p. 22). In other words, assessment gives feedback on students' progress while also informing teachers about the quality of instruction they are providing. Language assessment, according to Malone (2013), is a crucial and integrated element of education, and its function in the teaching promotes students' language development. In addition, Stiggins (2006) proposed that assessment is not something that occurs at the conclusion of the training but rather is an educational intervention that takes place several times all throughout the instruction. In this sense, the primary goal of assessment is not to determine the level of mastery attained by students as a direct result of teaching; rather, it is to identify areas in which teaching and assessment can be improved, given that these aspects are the basis of the teaching and learning process (Deluca, Luu, Sun, & Klinger, 2012; Stiggins, 2007). Therefore, as a result of assessment practices, not only teachers but also students receive ongoing feedback informing about students' development (McDowell et al., 2006). Based on these explanations, assessment is an essential part of teaching, and learning.

2.9. Basic Concepts in Assessment: Assessment, Measurement, and Evaluation

Many individuals may find the phrases assessment and testing to be confusing. According to Katz (2012), testing is merely one kind of assessment, and assessment refers to a multitude of methods and processes for gathering data to make judgments on learning. In a similar vein, Brown (2003) defined assessment as the umbrella term that covers both tests and tasks.

The term "assessment" in educational sciences has been given a wide variety of meanings, and various researchers working in the fields of educational measurement and language assessment have used this term in a variety of different ways, which suggests that there is no consensus concerning what exactly it means (Bachman, 2014, p. 7). In addition, "test(ing)," "measurement," and "evaluation" are frequently used interchangeably for talking about assessment. Nonetheless, they have distinct meanings and do not refer to the same notion. According to Palomba and Banta (1999), assessment refers to "the systematic collection, review, and use of information about educational programs undertaken to improve learning and development" (p. 4). Norris (2006) describes assessment as the systematic collection of student learning-related data. According to Chan (2008, p.7) "assessment refers to any method, strategy or tool a teacher may use to collect evidence about students' progress toward the achievement of established goals". Evaluation is the overarching phrase for assessment and testing. Evaluation, according to Bachman (1990), is "the systematic gathering of information to make decisions" (p. 21). It has to do with making decisions and value judgments. Evaluation considers a wide variety of information sources; assessment, on the other hand, is a narrower phrase that focuses primarily on student learning. Measurement is another essential term in assessment. According to Bachman (1990), it is the "process of quantifying the characteristics of an object of interest according to explicit rules and procedures" (p. 18). Similarly to assessment, measurement may also refer to the product or result of the measuring process. Measurement is a form of assessment including quantification, or the assigning of numbers.

2.10. Types of Assessment

There are different kinds of assessments including formative and summative assessment, informal and formal assessment, direct and indirect assessment, objective and subjective assessment, discrete point assessment and integrative assessment, and finally norm-referenced assessment and criterion-referenced assessment.

2.10.1 Summative Assessment

The two major types of classroom assessment are formative assessment and summative assessment (Brown, 2003). The terms assessment of learning and assessment for learning are also used to describe summative assessment and formative assessment. They are primarily concerned with the interpretation of assessment results depending on the times at which the assessments are administered and for what purposes the findings are employed (Carr, 2011).

To begin with, summative assessment focuses on student success and is administered after the learning process has concluded to evaluate learning process and make decisions about student learning based on the grades given (Popham, 2009; Semerci, 2015). When an exam is administered at the end of a unit, program, course, etc. to determine how well students have mastered the subject, it is known as a summative test, and it is frequently used for grading reasons. Therefore, summative assessment does not occur during the learning process, nor does it alter learning; rather, it serves as a final evaluation of learning (Brown, 2003). Therefore, summative assessment is comprised of final examinations, as well as end-of-module /semester/ year exams and these assessments are designed to evaluate the knowledge and abilities that students have obtained after the completion of the learning process (Cheng, Rogers & Hu, 2004; Taras, 2005).

2.10.2. Formative Assessment

Formative assessment is the sort of assessment provided to students while they are still in the learning process to give information on the quality of their learning that is occurring (Bachman, 1990). Formative assessment is the practice of assessing and evaluating students throughout the teaching process in order to continually examine their progress and discover their weaknesses and needs (Semerci, 2015). In order to improve students' ability to learn and to uncover and minimize any learning problems that may exist for them, the formative assessment is carried out. In most cases, the information that is gained through formative assessment is utilized to assist in making judgments on whether or not there is a need for changes to be made to the course curriculum, instructional strategies, program, etc. Therefore, this kind of evaluation need to be seen as an integral component of the learning process. The goal of formative assessment is to provide guidance for both learning and teaching (Harlen, 2005). To put it another way, thanks to formative assessment, teachers can obtain feedback from their students, which enables them to identify weaknesses in both their own teaching and in the learning of students and design an action plan to compensate for these deficiencies in an appropriate manner. The achievement of students should not be determined for grading or other reasons based on the findings produced. Improving the effectiveness of teaching and learning is the primary objective of this assessment

2.11. Principles of Language Assessment

In order to construct a quality and defensible test, certain criteria such as validity, reliability, authenticity, practicality and washback should necessarily be taken into consideration (Farhady, 2012). These principles are explained in the following sections.

2.11.1. Validity and Reliability

Validity is one of the most essential components of assessment since it influences assessment quality. Validity has traditionally been described as the "extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment" (Gronlund, 1998, p. 226). Henning (1987, p. 89) described validity as "appropriateness of a given test or any of its components as a measure of what it is purported to measure". That is, validity based on to what extent a test measures what it is supposed to measure (Hughes, 2003). For Gronlund (1998, p. 226), validity is "the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment". Validity can be classified into four main categories (Harrison, 1983); content validity, construct validity, criterion-related (or empirical) validity, and face validity.

To begin, content validity refers to the degree to which a test accurately portrays the content of the exam in terms of validity (Henning, 1987). According to Wood (1993), when topic selection is not done effectively in terms of validity, the test results may not reflect the truth. According to Akbari (2012), a test lacks content validity if it fails to adequately reflect the domain it is intended to measure. Therefore, content selection is extremely important for a test's content validity (Henning, 1987). For content validity, what should be done is a rigorous and extensive examination of the relevant skills and structures, and then the items on the test should be representative of all of the skills and structures that are covered (Öztürk & Aydın, 2018).

In order to comprehend construct validity in connection testing, it is essential to first comprehend the notion of construct. Despite its potentially deceptive name, it has nothing to do with the design or construction of an exam. It is possible to define construct as "the ability, skill, or proficiency that happens in the human brain" (Shiken, 2000, p.9). To illustrate construct validity refers to the extent to which a test assesses the intended construct (Ying & Wei, 2016).

When it comes to criterion-oriented validity, the objective of criterion-oriented validity is to determine whether there is a connection between the scores of two comparable tests that assess the same construct (Akbari, 2018). Criterion-oriented validity stresses the degree to which the findings of a test coincide with the results of another independent and trustworthy assessment in which the test taker participated for an ability (Brown, 2003). If two tests assess the same construct, there is expected to be a correlation between their findings. In the event that this is not the case, it indicates that one of these tests has a poor design or assesses a different construct.

Regarding face validity, it refers to how much test-takers believe the assessment is accurate, fair, appropriate, and beneficial for enhancing learning (Gronlund, 1998). Face validity is primarily concerned with test-takers' impressions of whether or not questions are relevant to the domain (Abkari, 2018). In other words, face validity focuses on whether test-takers consider the physical characteristics of the questions, such as the structure or layout, to be meaningful. Henning (1987) contends that asking test-takers if the exam fulfilled their expectations can assist to evaluate face validity since test-takers' perceptions impact face validity. Consequently, if test-takers say that the exam they took fell short of their expectations, it may be concluded that the test lacks face validity.

A test's reliability is determined by its consistency and accountability (Brown, 2003). In this respect, a test is deemed reliable when it produces consistent results at different times of administration (Hughes, 2003). In other words, if the same exam is conducted to a group of students in a variety of test scenarios, there should not be any change in the students' scores across context of the test and occasions. Student-related reliability, rater reliability, test administration reliability, and test reliability are elements that influence reliability (Brown, 2003).

A test's reliability may be impacted by a variety of rater-related variables, including error, subjectivity, prejudice, disregard for the scoring criteria, and inexperience. Rater reliability is divided into two as intra-rater reliability and inter-rater reliability. Inter-rater reliability exists when the scores offered by two or more teachers for the same exam are consistent and similar (Harris, 1969). Intra-rater reliability arises when the scores given by the same rater are consistent. Low intra-rater reliability may stem from a variety of causes, such as unclear scoring criteria, insufficient expertise or attention, tiredness, etc (Brown, 2003). When compared to multiple-choice tests, free-response tests have a tendency to have lower rater dependability (Harris, 1969).

The student-related variables that influence a test's reliability include physical or psychological factors including fatigue, demotivation, and anxiety that students experience on the day of the assessment (Brown, 2003).

The reliability of test administration relates to the physical conditions under which tests are conducted, such as lighting, noise, temperature, the quality of copies, and seating arrangements (Brown, 2003). Test reliability is also related to test being applied to students and it has the power of the quality of reliability. The design of tests that are too long or too short or have unclear instructions makes them unreliable because test-takers may feel overburdened by the time commitment or confused by the test's complexity (Brown, 2003).

2.11.2. Practicality, Authenticity and Washback

The term "practicality" describes how useful, practical, and economical a test is (Cohen & Swerdlik, 2013). Tekin (2000) also defines practicality as the test's ease of preparation, replication, application, and scoring. According to Bachman and Palmer (1996), the design, preparation, and implementation of a practical test should not need the use of more resources than are now accessible. When a test requires more time to complete than expected or is too costly to take, it is deemed impractical (Brown, 2003).

Authenticity is another crucial factor that must be incorporated into the test design. Bachman and Palmer (1979) defined authenticity as "the degree of correspondence of the characteristics of a given language test task to the features of a target language task" (p.23). The use of genuine daily language, meaningful topics, and the selection of contextualized and real-world tasks enhances the authenticity of a test (Brown, 2003). According to Brown (2001), an assessment is authentic if the language of the test items is natural, the items are context-based, and the activities include real-life scenarios. Furthermore, the exercises should be relevant to daily life and engaging to the students.

Washback is one of the components of assessment, and there are several definitions of it in the literature (Bachman & Palmer, 1996; Carr, 2011; Cohen, 1994; Hughes, 2003). Washback is defined by Hughes (1989:1) as "the effect of teaching and testing on learning." Testing, he claims, may have a beneficial or negative impact on students. Washback is defined by Cohen (1994) as a way assessment tools influence instructional practices and beliefs. In a similar manner, Carr (2011) described it as the effect that a test has on both the curriculum and the way that students prepare for exams. In other words, it is the influence that a test has on teaching and learning. Exams have apparent influences on student motivation,

classroom activities, school, the education system, and society in general (Wall, 2005). Therefore, it can be suggested that washback may influence students, test developers, and the education system. (Bachman & Palmer, 1979).

To summarize, the criteria listed above are critical for assessment practices. It is crucial that language teachers, test developers, and administrators consider these principles while assessing students' abilities.

2.12. Assessment Literacy

Assessment is an integral part of the teaching and learning process, and instructors play a significant role in assessment. Teachers have several responsibilities in assessment-related tasks. To begin, Webb (2002) defined assessment literacy as teachers' understanding of how to assess their students' development, as well as how to interpret and utilize assessment data to promote learning and boost instruction. In a similar vein, Davies (2008) described it as the combination of abilities in constructing and evaluating a test, as well as knowledge of assessment evaluation, measurement. In this respect, instructors must have enough understanding of the major concepts of assessment and be aware of their effects on the design of assessment procedures and actions taken based on assessment results (Rogler, 2014). According to Stiggins (1991), assessment literacy is the capacity to discriminate between low-quality and high-quality assessments and draw conclusions about student success based on this information. It is viewed as a link between learner accomplishment and assessment quality (Mertler, 2009). For Falsgraf (2005, p. 6), assessment literacy is “the ability to understand, analyze and apply information on student performance to improve instruction”. Based on these definitions, it is reasonable to conclude that instructors are at the heart of assessment, and they are required to possess particular abilities in order to carry out assessment successfully. In other words, teachers can be considered language assessment agents (Rea-Dickins, 2004). According to Mertler (2009), teachers should be aware of the wide range of assessment-related activities, as well as their strengths and weaknesses, and this awareness is a requirement for successful and effective assessment. As a result, teachers are expected to identify learning objectives and develop assessment activities accordingly in order to fully understand students' progress. In conclusion, it is vital for teachers to be able to develop and choose an effective assessment tool or technique that aligns with the program's objectives, to be aware about assessment principles and techniques, and to be proficient at implementing them in classroom assessments and using assessment findings to improve

education (Popham, 2004). Teachers with a high degree of assessment literacy are able to draw conclusions from assessment data, better understand students, and develop long-term learning goals (Gottheiner & Siegel, 2012). Teachers and educators must therefore have detailed knowledge of assessment. Evidently, assessment literacy is an essential competence for teachers to have. However, Brookhart (2001) demonstrates that teachers have inadequate assessment skills. Unfortunately, they possess an inadequate level of assessment literacy. According to Stiggins (2001), a lack of assessment literacy results in inaccurate assessment findings; hence, students are unable to reach their full potential. Also, this situation makes it challenging for teachers to perform assessment tasks correctly (Mertler and Campbell, 2005). That's why teachers should be assessment literate to "support and improve both learning and teaching" (Haught & Crusan, 2016, p. 179).

2.12.1. Studies on Assessment Literacy and Assessment Knowledge of Teachers

It is obvious from the existing literature that teachers do not have sufficient assessment literacy. Measuring assessment literacy levels of the teachers to detect their strengths and weaknesses, Impara, Plake and Fager (1993) carried out a study employing Teacher Assessment Literacy Questionnaire developed based on the "Standards for Teacher Competence in Educational Assessment of Students" (AFT, NCME, & NEA, 1990). As a result, in-service teachers were found to have low level of assessment literacy.

Mertler (2003) investigated instructors' assessment literacy levels using the Classroom Assessment Literacy Inventory. Both pre-service and in-service teachers participated in this study. According to the findings, in-service teachers outperformed pre-service teachers in terms of assessment literacy.

Mertler and Campbell (2005) developed a second tool to measure teachers' assessment literacy. This tool was also associated with "Standards for Teacher Competence in Educational Assessment of Students" (AFT, NCME, & NEA, 1990). Responses to scenario-based questions revealed that pre-service teachers in this research have a poor level of assessment literacy.

Volante and Fazio (2007) conducted a study with 69 pre-service teachers from each of the four years of the ELT program using a different research methodology. A survey with four closed and five open-ended questions was administered to the participants. The data revealed that participants' self-efficacy ratings were extremely low over all four years of the program.

Furthermore, the pre-service teachers expressed an urgent need for a specialized course focused on classroom evaluation.

Mertler (2009) examined the effect of a two-week training based on "The Standards for Teacher Competence in the Educational Assessment of Students" on the assessment literacy of in-service teachers. Assessment Literacy Inventory was used prior to and after the workshop. Additionally, reflective journals were employed to gather in-depth data regarding the experiences of the teachers. The findings demonstrated that instructors scored better on the post-test and had a favorable attitude on the development of assessment literacy. As a result, training was viewed as having a good impact on instructors' assessment literacy.

Tao (2014) conducted a research with in-service EFL instructors using four distinct measures that focused on classroom assessment knowledge, innovative methods, grading bias, and quality procedure, respectively. The first one is designed to examine instructors' assessment knowledge. The purpose of the remaining three measures was to determine the participants' assessment-related beliefs. In addition to the quantitative data, semi-structured interviews with six teachers were held. The results indicated that the teachers had insufficient knowledge about assessment.

Xu and Brown (2017) also conducted research to learn more about English instructors' assessment literacy in China. The findings revealed that the majority of participants had a basic or low level of assessment literacy. It was also demonstrated that teachers' demographic characteristics such as age, years of experience, title, and assessment training experience had no significant influence on their assessment literacy.

Based on the findings listed above, we may conclude that both pre-service and in-service instructors have limited knowledge about assessment. According to Stiggins (2007), instructors will need to improve their assessment literacy levels in the future. They were not previously expected to be more assessment literate, but changing times have shifted general perception about teachers, and there is growing demand on them to be much more assessment literate.

2.13. Language Assessment Literacy

Although the phrase "language assessment literacy" first used in regard to AL, it now refers to a field independent from AL. In other words, Inbar-Lourie (2017) states that the word LAL is derived from AL, however LAL differs from AL in that LAL attempts to set

itself apart as a knowledge base that incorporates unique aspects inherent in theorizing and assessing language related performance" (p. 259). Several definitions of LAL have been proposed in the literature. First of all, it should be noted that LAL demands additional abilities compared to assessment literacy, and LAL is the combination of assessment literacy skills and language-specific skills (Inbar-Lourie, 2017). First, according to Malone (2013), LAL refers to "language teachers' familiarity with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing language" (p. 329). Lam (2015) also defined it as "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). Additionally, language assessment literacy was defined by Taylor (2009) 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).

LAL has a great importance for language teachers. Scarino (2013) stated that LAL is a necessary skill for language teachers because thanks to LAL, language teachers can "explore and evaluate their own preconceptions, understand the interpretive nature of the phenomenon of assessment and become increasingly aware of their own dynamic framework of knowledge, understanding, practices and values (p. 311). According to Malone (2013), assessment and teaching complement and strengthen one another. In other words, instructors who are assessment literate are able to interpret test scores in order to discover the areas in which their students struggle and, as a result, improve both the quality of their instruction and the education their students get (Mertler, 2003). However, despite the importance of assessment in the teaching and learning process, assessment training for language teachers is inadequate (Lam, 2015). Briefly, the research reveals that language teachers must have assessment literacy in order to perform successful assessment practices that are aligned with learning objectives and assessment goals.

2.13.1. Studies on Language Assessment Literacy and Language Assessment Knowledge of Teachers

Language assessment literacy research has focused on a variety of yet related topics, including teachers' assessment requirements and perspectives, pre-service assessment courses, and teachers' assessment knowledge levels. Several studies have examined the present levels of LAL among EFL teachers in various contexts. Tsagari and Vogt (2017), for example,

found that participants in their study, who were teachers from Cyprus (n=16), Greece (n=22), and Germany (n=25) without assessment training, felt themselves insufficient in terms of LAL.

In addition, Volante & Fazio (2007) conducted another study with pre-service teachers, and they discovered that the participants self-rated themselves having poor LAL levels. It was also discovered that the participants utilized testing and assessment tools for traditional summative purposes. In another study, Sultana (2019) interviewed ten language instructors in Bangladesh to explore their assessment literacy, and the findings revealed that the participants' awareness of assessment practices was inadequate. Investigating teachers' language assessment beliefs and practices, Zhang and Burry-Stock (2003) found that teachers utilize various assessment techniques based on classroom level. While secondary teachers prefer objective assessment techniques, primary teachers prefer alternatives.

Muoz et al. (2012) used surveys, interviews, and a written report of experiences to find out how EFL teachers thought about and used oral and written assessments. The results showed that what teachers think and what they do about language assessment don't match up. Even though teachers think that assessment practices could help them teach and learn better, they don't usually use effective assessment techniques in their practices accordingly. In addition, Jannati (2015) conducted a qualitative study with 18 Iranian EFL instructors to determine their perspectives and evaluation techniques. Analysis of the qualitative data collected through semi-structured interviews revealed that regardless of their experiences, all participants possessed an equivalent degree of assessment knowledge. Finally, it was observed that assessment-literate teachers' practices do not match their perception.

Hakim (2015) also examined the assessment literacy of thirty Saudi Arabian EFL instructors. Participants were given a questionnaire on teachers' assessment ideology to find out their beliefs of language assessment literacy and to determine if teaching experience influences teachers' perceptions of language assessment knowledge and practices. The findings demonstrated that EFL instructors were knowledgeable of main language assessment techniques, and that they effectively put their knowledge into their practices. Regarding teaching experience, although both novice and experienced instructors appeared to have substantial understanding of language assessment techniques, those with more experience were found to implement them more effectively in the classroom.

In accordance with international research, studies conducted in Türkiye indicated a need for language assessment literacy training (Mede & Atay, 2017). Firstly, Hatipoğlu (2015) also

conducted research into the knowledge and needs of pre-service language instructors. This mixed-methods study included both a survey and interviews. The quantitative data revealed that the assessment literacy of future language teachers was poor. The interviews also revealed that pre-service teachers believed a single course in language assessment is insufficient to qualify as assessment literate.

Furthermore, Büyükkarcı (2016) investigated the LAL levels of thirty-two EFL teachers from elementary schools, high schools, and university preparation programs. It was discovered that participants, regardless of their teaching context, had a very poor level of assessment literacy and required more training in the sector. Mede and Atay (2017) examined the evaluation literacy of 350 Turkish EFL instructors employed in preparatory programs at both foundation and state universities. There was a need for training in classroom-based assessments and assessment-related concepts and principles, as demonstrated by the participants' low assessment literacy. Participants did not feel competent in assessing productive and receptive skills, but they felt competent in assessing grammar and vocabulary, according to the qualitative findings of their study.

Ölmezer-Öztürk and Aydın (2018) conducted a study in English preparation programs in Türkiye about the assessment knowledge of EFL teachers. The researcher administered the Language Assessment Knowledge Scale (LAKS) to 542 EFL teachers after designing and verifying it. In addition to the scale, 11 participants were interviewed regarding their language assessment knowledge. The participants' language assessment knowledge was found to be inadequate. It was also shown that teachers were most knowledgeable about assessing reading and least knowledgeable about assessing listening. Teachers indicated that their pre-service and in-service trainings were insufficient and that they need assessment training. Hatipoğlu (2015) surveyed 124 pre-service EFL instructors in an ELT department to determine their understanding of language assessment and their expectations for the English Language Testing and Evaluation course. Interviews and needs analysis survey questionnaires were used to obtain data. The majority of pre-service teachers anticipated preparing and evaluating English language tests for national and international language examinations, despite having limited expertise of testing and assessment, according to the results. As a result, it was stated that a lack of assessment literacy among English language instructors in Türkiye may result in ineffective assessment implementation, therefore testing and assessment training programs should be given greater priority.

Previous research has demonstrated that Turkish EFL teachers assess and evaluate their students' language abilities without sufficient knowledge of assessment. Several factors, including a lack of training and standards, the curriculum, and the workload, may have a negative impact on their assessment processes. However, as shown in the literature, assessment literacy is important in education; consequently, it requires further investigation.

2.14. Theoretical Framework

The theory underlying this study is sociocultural theory (SCT). This part firstly provides information about constructivism and social constructivism and their explanations related to learning and teaching. Next, it follows with claims of SCT on teacher learning.

Constructivism departs from the idea that there is no objective reality representing knowledge; instead, knowledge is created through adaptive cognitive processes such as assimilation and adaptation, resulting in individuals' schemata being modified to include new information (Piaget, 1972; William & Burden, 1997). The central concept of constructivist learning theory is that learners create their own knowledge by associating new information with the existing one. This theory posits that people have unique ways of experiencing reality and describes perception as a mental process in which learners construct their meaning and knowledge (Gray, 1997). Other researchers, such as Caffarella and Merriam (1999), agreed that learning is a process of constructing meaning. Furthermore, Duckworth (1987) clarified that humans do not receive but rather provide or attach meaning to events around them to state that individuals create meaning. Constructivism's primary goal is restructuring rather than disseminating information, and it calls for individual participation, inquiry, problem-solving, and cooperation (Abdal-Haqq, 1998; Akinoglu, 2013).

When comes to social constructivism, as the name implies, it emphasises the importance of interpersonal engagement in a society (socio) and the use of cultural-historical artefacts (cultural) as fundamental components of learning (Vygotsky, 1978). This perspective usually posits that learning and teaching occur as a result of social interactions in a social setting (Lantolf, 2006). Also, the SCT claims that social interactions facilitate individual development; group members exchange cultural meanings, which are finally internalised by people (Ilin, 2003).

According to this viewpoint, knowledge construction is a socio-culturally mediated process influenced by physical and psychological factors such as artefacts, parents, students, teachers, and the context. That is to say, knowledge cannot be an individual construction as it

results from a shared interaction (Doolittle, 2003). Johnson (2009) defines learning as a “progressive movement from external, socially mediated activity to internal mediation control by individual learners.” (p.17). According to Vygotsky (1978), humans acquire and internalise concepts through an initial interpersonal developmental process. This process stresses language as a primary instrument when concepts are being learned from others and internalised through private speech, which enables individuals to self-regulate their learning. That is to say, people co-construct knowledge in social environments and use language to give events meaning. According to Vygotsky (1978), the transition from interpersonal to intrapersonal processes is critical in developing higher mental processes.

According to this sociocultural perspective, internalisation, mediation, zone of proximal development, and scaffolding are the four basic concepts that make up human learning as a social practice. First, internalisation is a process that allows individuals to interact with their physical environment and turn all of their gains into psychological and mental reality. According to Vygotsky (1978), this process consists of two levels: the social level occurring among people through interaction and the individual level, which takes place inside the individual. People reconstruct their relationship with the environment through this process, which is always socially mediated regardless of the actual presence of other people (Winegar, 1997). Mediation is the second tenet of the sociocultural perspective. It is based on the idea that cognitive functions are mediated by individuals' interactions with their surroundings (Lantolf & Thorne, 2007). The next principle of the socio-cultural view is the zone of proximal development (ZPD). ZPD is the area of competence that exists between the area of competence that a person can learn on their own and the area of competence that an individual can only attain with the assistance of an adult, typically a more knowledgeable other (Abtahi, 2017). According to this perspective, ZPD is a tool for comprehending the internal mechanisms people employ when learning and addressing problems. As a result, ZPD is a significant construct for the purpose of concentrating on the developmental processes of adult learning. Scaffolding, the fourth principle of sociocultural perspective, is closely associated with ZPD. Scaffolding may be defined as an aid given by a more experienced person to a less experienced one in order to foster collaboration.

Constructivism is often linked to how students learn (Wang, 2011), but teachers also use it to create new knowledge (Shermila, 2013). Students gain knowledge by using inquiry and problem-solving techniques to come to conclusions and develop educated predictions, both by themselves and in cooperation with their peers (Shermila, 2013). The same can be true of

teachers, who cooperate with their colleagues to reflect on and understand several topics, including test materials, syllabi, student resources, and many more (Ostovar-Nameghi, & Sheikahmadi, 2016). Constructivism is often regarded as an advantageous tool in teacher education because of various aspects, including providing space for creation, allowing flexibility to address concerns, and promoting the internalisation of learning (Abdal- Haqq, 1998). Bishop (2001) explains that teachers also apply constructivism since they continue learning from past experiences and they also re-evaluate their existing knowledge upon each new interaction and reach a new degree of understanding (Driscoll, 2000).

The sociocultural perspective incorporates both theoretical and practical implications for teacher education and serves as a road map for better instructional practices. This theory emphasises, for example, that learning and teaching are joint activities that are affected by interactions among individuals in social environments to foster intellectual development (Larsen-Freeman, 2000). The socio-cultural approach is appropriate for tracing teachers' professional growth since it considers historical elements such as personal experience and formal development activities (Feryok, 2012). It is, therefore, not surprising that teachers' professional development has evolved in recent years towards a social constructivist perspective (Pitsoe & Maila, 2012).

According to the social constructivist theory, which contends that social interaction, mediation, and reflection are key components in the construction of knowledge, teachers should have the chance to participate in activities that require verbal interaction and communication with both novices and experts (Rock & Wilson, 2005). Engaging in these activities within a community of practice facilitates teacher development by fostering collegiality by sharing ideas. In this way, the sociocultural perspective combines social and cognitive components of learning. These initiatives emphasise giving teachers opportunities to comprehend theories in their workplaces (Freeman & Johnson, 2005). Von Glasersfeld (1989) also claimed that teachers actively construct knowledge in their educational surroundings through interaction and negotiation of meaning rather than passively acquiring it. In light of this, teacher professional development has shifted toward a social constructivist viewpoint by concentrating on their zone of proximal development (Pitsoe & Maila, 2012). According to the sociocultural view, teacher development is perceived as situated learning incorporating scientific and experienced discourse. Correspondingly, the construct called the zone of proximal teacher development meaning “the distance between what teaching candidates can do on their own without assistance and a proximal level they might attain through

strategically mediated assistance from more capable others (i.e., methods instructor or supervisor)” (Warford, 2010, p. 253) emerged. The basic insight of this strategy is the idea that learning leads to progress.

Recent research on second language teacher education, specifically on teacher cognition, has evolved from positivist to sociocultural approaches to teacher learning. SLTE formerly focused on how languages were acquired. However, now it investigates how instructors come to know what they know, how ideas develop in time, and how learning processes have developed them. Understanding how the mental processes of teachers evolve is crucial to this theory. The sociocultural perspective can provide a substantial theoretical framework for language teacher cognition research. According to Johnson (2009), social and cultural activities also shape human cognition, a crucial concept for sociocultural theory. According to him (2009), the social activities, in which a teacher engages, affect his or her cognition. In other words, instructors' experiences significantly shape their knowledge and ideas. Cross (2010) develops a framework for language teacher cognition using a sociocultural perspective. By characterising language instructors as agents within social settings, he argues that the sociocultural perspective aids researchers in unearthing their historical, sociological, and cognitive conceptions. He proposes that a sociocultural framework improves understanding of how teachers know, think, and conduct. Cross (2010) provides the following overview to highlight the link between sociocultural perspective and language teacher cognition.

A sociocultural theoretical perspective on teacher practice provides the basis for a systematic, comprehensive, and theoretically robust framework that accounts for the social dimension of thought and knowledge (p.449).

As a result, a sociocultural perspective may be utilised to understand better how teachers' social and mental activities are organised via culturally and socially produced relationships (Lantolf, 2001).

CHAPTER III

MEHTOD

3.1. Introduction

In this chapter, research paradigm and design of the study are discussed first, followed by the participants and sampling procedure. Second, an explanation of the setting of the study is provided. Following this, the RRD session processes are explained on a session-by-session basis. Thirdly, information on the objectives and content of each data collecting instrument, as well as how and when each tool is deployed during the research, is provided. Finally, the data analysis procedures, research trustworthiness, ethical issues, researcher's role are explicated.

3.2. The Research Paradigm and the Research Design

There are two dominant research paradigms in social sciences such as positivist inquiry and naturalistic inquiry. Each paradigm has a distinct nature as well as its own specific characteristics. It is important to note that both types of paradigm offer unique ways for interpreting the data that is gathered (Creswell, 1994). Although it is criticized for neglecting both individuality and uniqueness and for thinking that external facts of the world may be explored and understood by the researcher, positivism focuses on observable facts with an emphasis on explanation and prediction (Cohen & Manion, 2007; Maykut & Morehouse, 1994). In contrast, naturalistic inquiry is more concerned with comprehension and exploration. Therefore, it is possible to indicate that naturalistic inquiry is a prevalent method of research in the social sciences.

Moreover, while addressing positivist and naturalist paradigms, the phenomena of objectivity and subjectivity frequently come up. It is regarded as subjective since the naturalistic research paradigm considers any event holistically. Maykut and Morehouse (1994) have coined the word 'perspectival' in light of the fact that the term 'subjective' has certain negative connotations that lead naturalistic investigation to become notorious. These two distinct research methodologies generate distinct sorts of data and, thus, different outcomes. It should be noted, however, that these two sorts of data are seen as complementing rather than opposed to one another (Fitzpatrick & Meredith, 2006).

Researchers pursuing knowledge are said to be guided in their actions by a research paradigm. In fact, what is intended by the acts is the research itself. It is critical for a

researcher to properly identify the research paradigm since it determines the quality of study. Ontological and epistemological assumptions should inform all studies. In other words, some beliefs about the nature of knowledge and reality lead to the emergence of particular beliefs about knowledge production, which in turn have significant consequences for determining research design (Buberley, Johnson, & Cassel, 2012). As a result, determining whether to use qualitative or quantitative technique for a study is influenced by epistemological and ontological beliefs. (Hatch, 2002). That is, based on these reasons, it may be claimed that in order to establish a well-developed research framework, a researcher's epistemological, ontological, and methodological viewpoints should be included into the accepted paradigm.

First, a paradigm may be seen as a collection of fundamental assumptions about how individuals make sense of the world and how they control their actions. (Guba & Lincoln, 1994). A paradigm is a common worldview that represents views and values in a particular field, according to Schwandt (2001). Guba and Lincoln (1994) propose that some specific questions might be asked in order to better comprehend the paradigms such as for epistemology (how do we know what we know? – what counts as knowledge), for ontology (what do we believe about the nature of the reality? – nature of reality), for axiology (what do we believe is true? –role of values in research) and for methodology (how should we study the world).

This study was designed as a case study and it adopts social constructivist (i.e., interpretive) paradigm. Accordingly, it may be claimed that this study is driven by the ontological premise that there is no single reality but rather that reality is socially built and that people form their own particular realities via interactions with others (Chilisa & Kawulich, 2012; Creswell, 2013). This present study aims at gaining insight into development of EFL instructors' language assessment cognitions. Furthermore, the social constructivist paradigm necessitates a constructivist epistemology, which states that knowledge is generated by humans, that is participants and researchers, as a result of their subjective interpretations and experiences. Based on this, it may be stated that the knowledge base is made up of the participants' subjective interpretations. As a result, the researcher should establish a close relationship with the participants, on which subjective evidence is founded (Creswell, 2013). It is believed that triangulation of data through semi-structured interviews, non-participant classroom observations with field notes, KWL charts, reflection reports served this purpose in the current study. Views, perceptions, and subjective experiences of both participants and researchers are incorporated into the concept of multiple realities in the present study. In

addition, it is understood that values are distinct and that there is no clear boundary between good and wrong. In other words, each group has its own set of values, and it cannot be said that any of them are absolutely wrong or correct. In light of this, it may be argued that social inquiry is value-laden and that researchers should admit the nature of the study and convey their biases and values (axiological assumptions) in a clear and concise manner (Chilisa and Kawulich, 2012). The data acquired from each participant in the current study is equally valuable and cannot be generalized to the universe or other situations. Furthermore, it should be noted that beliefs and values have played an important part in the inquiry since the beginning. Regarding the methodological assumptions of the interpretative paradigm, this dissertation was structured as a case study, in which a researcher investigates an issue from the perspectives of the participants in a particular context, using multiple sources of data collection and getting close to the participants in their natural environments (Creswell, 2013; Yin, 2014).

Because each research paradigm encourages its own research design, the premises of the interpretative (constructivist) paradigm correspond to qualitative research methodology in order to identify and understand the meaning (Creswell, 2014). It also requires that the study be performed in the participants' own environment. Furthermore, it is necessary to gather data by incorporating multiple types of data collecting instruments into the data collection process, which includes talking to participants and seeing them in their normal circumstances when they are behaving naturally rather than putting the participants in labs (Creswell, 2014). Considering the aforementioned accounts, this dissertation was carried out as a qualitative research. It is defined by McMillan and Schumacher (1993) as “primarily an inductive process of organizing data into categories and identifying patterns among categories” (p.479). In addition, Creswell (1998) describes qualitative research as an inquiry process based on various methodological traditions that investigates a social or human issue and the study constructs a sophisticated, holistic picture, analyses language, reports specific opinions of informants, and takes place in a natural context. Creswell (2007) present key characteristics of qualitative research including “natural setting, researcher as key instrument, multiple sources of data, inductive data analysis, participants’ meanings, emergent design, theoretical lens, interpretive inquiry, and holistic account (pp.37-39). In addition to its benefits, the qualitative research approach has been criticized for its disadvantages in terms of sample size, generalizability, researcher role, time used, and hard labor required (Dornyei, 2007). Despite of these criticisms, qualitative research methods have been extensively utilized as a manner of

inquiry in social sciences (Marshall & Rossman, 1999). In order to understand how individuals construct the world around them and to fully explore a phenomena in order to provide explanations, qualitative research methods are particularly useful. Also, qualitative research approaches assist researchers to obtain a deeper understanding of any social phenomena in a participant's natural environment (Bryman, 2014), in addition to offering thorough and elaborate information about the investigated issue (Schmidt, 1983). They are priceless in terms of “describing multi-dimensional, complex interpersonal interaction where the limited focus of quantitative measures would be inadequate” (Krathwohl, 1998, p.243). Furthermore, the goal of a qualitative investigation is to discover new information rather than to validate a present notion (Sherman & Webb, 1988). In other words, the qualitative understanding helps researcher “seek to discover and understand a phenomenon, a process, or the perspectives and worldviews of the people involved” (p.11). Through qualitative research and extensive data collection, the topic of the current study was systematically and in-depth studied, resulting in a thick description with a sophisticated and thorough knowledge of the topic. It also increased the influence of participants' opinions by placing them at the center of the investigation. In addition, qualitative analysis was utilized not only for the purpose of unearthing the opinions of the participants, but also for the purpose of elucidating the factors that led to the participants holding such positions (Creswell, 2007). Consequently, a qualitative case study technique seems appropriate for giving a comprehensive description of the data in its real-world context and an explanation of the phenomenon's complexities (Zainal, 2007). Qualitative case studies are also advantageous because they allow the researcher to record or study the real world with all of its complexities in a way that is not possible with quantitative methodologies (Wallace and Atkins, 2012). Furthermore, Yin (1994) states that “in general, case study is a preferred strategy when questions such as how and why are being posed, when the investigator has little control over events, and when the focus is on a contemporary phenomenon within some real life contexts” (p.1).

In qualitative case studies, case is a basic unit of analysis. According to Patton (1987), depending on the context, cases might represent organizations, events, programs, or communities that are analyzed holistically. In a similar manner, case is defined as “a specific, complex and functional thing” (Stake, 1995, p.2). In addition to these, Miles and Huberman (1994) explains the case as “a phenomenon of some sort of occurring in a bounded context which is, in effect, the unit of analysis” (p.25). In addition, Creswell (2007) describes the case in a case study as a bounded system that must be studied in time in a holistic manner using

multiple types of data collecting instruments. Accordingly, in the current study, case is development of EFL instructors' cognitions of assessment and this case is bounded by space which is a school of foreign languages of a state university located in eastern part of Türkiye. Furthermore, it is bounded by time which is 2019 – 2020 academic year. From the Creswell(2007)'s definition of a case study which is an approach “in which the investigator explores a bounded system(a case) or multiple bounded systems(cases) over time through detailed, in-depth data collection involving multiple sources of information and reports a case description and case-based themes”(p.73), it could be understood that “if the phenomenon is not intrinsically bounded, it is not a case” which might be interpreted as case are bounded systems (Merriam,1998, p.27).

Case studies have been described by many researchers so far. To begin, Dörnyei (2007) described case studies as an ideal way for acquiring a detailed account of a complicated social issue embedded in a cultural context. It provides rich and in-depth insights that no other approach can provide, enabling researchers to analyse how a complex collection of events combine to shape the social environment around us. Yin (2003) states that case study is “an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used” (p.13). Last but not least, Creswell (2007) compiles different perspectives on case studies, analyses them, and then presents a more detailed explanation of case studies by stating that “case study research is a qualitative approach in which the investigator explores a bounded systems (cases) over time through detailed, in-depth data collection involving multiple sources of information (e.g. observations, interviews, audiovisual materials, documents and reports), and reports a case description and case-based themes” (p.73).

Baxter and Jack (2008) have proposed a series of actions to be taken while carrying out case studies including specifying the case(s), formulating research questions, determining the boundaries of case(s), deciding on the type of case study considering the criteria proposed by Yin (2014) and enhancing the triangulation and credibility by diversifying the data sources. As it is stated before, case for the current study is the development of EFL instructors' assessment cognitions. The case in this present study is bounded by space and time. In other words, case is bounded by a space which is a school of foreign languages of a university located in eastern part of the country. Also, the case is bounded by time due to limited data collection period-an academic year- and size of participants – 4 EFL instructors. As a

longitudinal case study, the present study is a collaborative study focusing on the developmental process of the participants' nature of current cognitions before RRD sessions as well as identifying their strengths along with weaknesses regarding assessment. Also, exploring development in participants' language assessment cognitions as a result of a set of RRD sessions, if there is any was targetted in the study. Considering these aims, following research questions were formulated for the current study as follows:

1. What is the nature of EFL instructors' cognitions regarding language assessment?
2. What are participants' strengths and weaknesses considering language assessment, as they perceive?
3. Do reflective reading and discussion sessions lead to any development in EFL teachers' cognitions concerned with language assessment, as they perceive? If so, what are the changes?

Later, type of case study was determined considering Yin (2003)'s classification of case study. Yin (2003) categorizes case studies as explanatory, exploratory, and descriptive and he makes differentiation between holistic single-case and multiple case studies. Considering the categories determined by Yin (2003), the current study adopts descriptive and holistic single case study types which "is used to describe an intervention or phenomenon and the real-life context in which it occurred" (p.24). This study employs a single case study design since it examines a single set of contextual conditions, as opposed to a multiple case study approach, which is utilized for analyzing situations from various and multiple contexts. When comes to the current study, the case was determined to be development of assessment teacher cognition in a school of foreign languages context. Also, in this study, researcher analyzed a single context from various perspectives. As for the data triangulation and credibility. When constructing a case study, there are a number of important factors to bear in mind to maintain its overall relevance. First, researchers must present enough information for readers to evaluate the validity of their study. Second, triangulation is a critical technique that enables the case to be evaluated from many perspectives. This data collection enhances data quality by validating findings (Knafi & Breitmsyer, 1989). Having separate researchers independently assess the data enhances the reliability of the analysis (Baxter & Jack, 2008). In the current study, the researcher strengthened and guaranteed the trustworthiness of qualitative data by analyzing their credibility, transferability, dependability, and

confirmability. To establish credibility, the researcher offered detailed descriptions of the participants, physical environment of the study, data collecting technique, and roles he took during the study. In terms of trustworthiness, several data collection strategies were used to improve the methodological triangulation (Cohen & Manion, 1994). Confirmability of the study was guaranteed by a comparison of the data. In addition, member checking was performed throughout the research, and data were coded and analyzed by two distinct coders in order to increase dependability.

In addition, it should be highlighted that case study research design is a successful strategy for longitudinally investigating a challenging issue (Van Lier, 2005). In such longitudinal case studies, data relevant to the objectives of the study are collected at various times in time. According to Menard (2002), in a longitudinal study, data is collected in two or more distinct time periods, and participants are examined either the same or comparably from one period to the next. Data analysis includes some comparisons of data between periods, and thus differences or changes in the examined case from one period to the next can be determined. Accordingly, the present study meets the requirements of a longitudinal case study, including selection of participants from the same population, use of comparisons across data sets and between periods, and collection of data over time. Therefore, development in participants' cognitions regarding assessment was examined and described in detail.

Based on the above explanations, it would be more accurate to state that careful research processes and triangulation were applied to establish trustworthiness in this qualitative case study, as the trustworthiness of positivist paradigm was not applicable.

3.3. Participants and the Sampling Method

This current research firstly aimed at finding out EFL instructors' current nature of their cognition regarding language assessment. It also tries to reveal development in teachers' cognitions considering language assessment, if there is any, as a result of RRD sessions. Firstly, research site of the study was chosen through convenience sampling (Cresswell, 2012). Cohen *et al.*, (2007) explain that convenience sampling "... involves choosing the nearest individuals to serve as respondents and continuing that process until the required sample size has been obtained or those who happen to be available and accessible at the time" (p.113-114). Also, Cresswell (2012) described convenience sampling as a method that enables researchers to select people who are willing and accessible to be examined. Secondly, the participants were chosen using criterion sampling. As Dörnyei (2007, p.126) indicated

“qualitative inquiry is not concerned with how representative the respondent sample is. Instead, the main goal of sampling is to find individuals who can provide rich and varied insights into the phenomenon under investigation so as to maximize what can we learn”. For this reason, the researcher carefully chose the cases to be included in the study based on his evaluation of participants’ ownership of the specific features being sought (Cohen *et al.*, 2007). In other words, during the selection of the participants, researcher had background information and used it in the sampling process so as to choose individuals who are believed to be “information rich” (Patton, 1990, p.169) collect the most suitable data (Frankel, Wallen, and Hyun, 2011).

In parallel with these statements, detailed criteria for the sampling were outlined below:

- They were required to currently teach full-time at an English preparatory program of a university
- They were required to be willing to participate in the study
- They were required to assess the students to whom they teach English because at some school of foreign languages, there is an assessment unit and it is responsible for most of the assessment procedure.

As a result of these processes, participants of this study included four EFL instructors working at an English preparatory program of a state university located in the eastern part of Türkiye. They were selected intentionally as a result of procedures aforementioned. They were believed to provide with rich data for the study. In the beginning, the study included six instructors. Before RRD sessions started, two of them had withdrew from the study due to their workload. Therefore, detailed explanation was given below considering four participants participating in the study from the beginning to the end. Also, while describing the participants, in detail confidentiality was considered such as using a pseudonym for each of the instructors.

P1 is a 42 years old female teacher who has been working for 14 years. She has been teaching English at the preparatory program for 4 years. Before starting to teach at tertiary level, she worked for MONE for ten years. She holds an MA in English Language and Literature. She gives integrated and writing courses in the program. During the data collection process, she was teaching 24 hours per a week. The participant did not have a specific responsibility related to language assessment. She was only assessing her students, to whom she taught English. The participant teacher stated that she graduated from the department of

English Language and Literature, and stated that she did not attend any training or workshop on language assessment, except for the assessment and evaluation in the education course that he took in her pre-service education years.

P2 is a female instructor with six years of experience. She holds an MA in Educational Sciences. Also, she is an MA student in the department of ELL. She gives 26 hours of integrated course per a week. The participant graduated from the faculty of education and did not receive any education or training related to language assessment other than the testing and assessment in language teaching course she took during her pre-service education years. In addition, In addition, she has not taken part in the testing unit so far and only evaluates students for whom she is responsible for teaching English.

P3 is a 36 years old male teacher who has been working in the program for 9 years. He has been teaching English for 12 years. He previously taught English in special institutions. He holds an MA in Educational Sciences. He is also a PhD candidate in the same field. During data collection process, he was teaching English to prep class students for 26 hours weekly. The participant also stated that he only evaluates his own students and does not have any educational background related to the subject, apart from the testing and assessment course he took at the undergraduate level.

P4 is a female teacher. She has been teaching English for 10 years and she spent all these years in the department she is currently working in. She holds PhD degree in ELT. She teaches 21 hours of English every week. She just took the testing and assessment course in her undergraduate years. Meanwhile, none of the participants has office duties in the program.

3.4. The Context of the Study

The present study was conducted at a school of foreign languages at a state university based in eastern part of Türkiye. It was founded in 2007. The school of foreign languages was founded in 2011. Students who are enrolled in English language and literature department are accepted to school of foreign languages. After registration, students are required to take a proficiency exam. If their grade is 70 or higher, they are exempt from the preparation class, and they are allowed to start ELL department. If they fail, they are required to take preparation class courses. The courses at school of foreign languages cover four main skills such as reading, writing, speaking and listening. In addition to these, integrated course through which four skills are taught integratively is given. According to Council of Higher Education (2008) the program lasts two semesters and students can only repeat the program

once. The weekly course hour of the program is 27 hours including 17 hours of integrated course, 2 hours of listening and speaking course, 3 hours of reading course and 3 hours of writing course respectively.

In this study, each participant's two hours of lessons were observed after each interview. Other than the classroom setting in which researcher observed participant teachers, RRD sessions were organized. These sessions were held after all participants had read the materials of the session. These sessions took place in the researcher's office. Sessions lasted for minimum an hour and maximum two hours depending on the topic chosen. The content of the reflective reading and discussion sessions was determined by researcher and his thesis supervisor at the end of pre RRD sessions phase of the study. Also, participants were consulted while determining the content and order of RRD sessions. As for the materials that teachers were supposed to read before each RRD session, they were determined collaboratively by the researcher himself, his thesis supervisor and the participant instructors. The reading materials were sent to the teachers before each session and they were asked to read all or some of them. During RRD sessions, teachers' discussions were audio recorded.

3.5. The Research Procedure

Designed as a case study, data collection procedure includes three stages. In other words, in order to achieve research objectives, the study was threefold including pre, while and post reflective reading and discussion sessions. Data for the current study were collected during 2019 – 2020 academic year and this intensive process lasted for almost six months. Before starting the study, the researcher got the permission from the ethics committee (Appendix A, Appendix H). Also, researcher took the permission of the participants via informed consent forms (Appendix G). Then data collection process started based on the schedule which was provided to the participants beforehand. All of the process was held in Turkish. However, participants preferred to fill in KWL charts and reflection reports in English.

The purposes of the current study are multiple. Firstly, it aims at revealing EFL instructors' existing cognitions regarding language assessment, and then finding out their strengths and weaknesses about the issue, if there is any. Based on participants' expressed needs, this study also aims at designing a set of RRD sessions. In addition, this study also seeks to answer to the question about whether proposed reflective reading and discussion sessions led any development on the participants' cognitions regarding language assessment.

To attain these goals, individual semi- structured interviews were held, classroom observations were organized, RRD sessions were conducted, KWL charts and researcher reflection reports were kept, follow-up individual interviews were also administered. By taking these aims into account, the current study was carried out in three stages such as first stage which lasted twenty-eight days (before the RRD sessions), second stage which lasted a hundred thirteen days (during RRD sessions) third / follow – up stage which lasted forty-one days (after RRD sessions). Table 1 summarizes the overall research procedure.

Table 1.

Overall Research Procedure

Phases of the Study and their lengths	Research Questions	Purposes	Data Collection Tools	Steps included in data collection procedure	Data Analysis Techniques
Pre-RRD sessions – twenty-eight days	What is the nature of EFL instructors’ cognitions regarding language assessment? What are participants’ strengths and weaknesses considering language assessment, as they perceive?	To reveal EFL instructors’ existing cognitions regarding language assessment, and then finding out their strengths and weaknesses about assessment, if there is any. To design a set of RRD sessions.	Pre-semi-structured interviews, classroom observation with field notes, post semi structured interviews and researcher reflective report.	Pre-semi structured interviews were administered to teachers. Teachers were then observed in their classroom Finally, post-semi-structured interviews were applied to teachers. Researcher kept reflective report throughout the session for each participant separately.	Inductive Content Analysis
During RRD sessions- a	Do reflective reading and	To see whether	KWL charts, researcher	Ten RRD sessions were designed based on	Inductive Content

hundred and thirteen days	discussion sessions lead to any development in EFL teachers' cognitions concerned with language assessment, as they perceive? If so, what are the changes?	proposed reflective reading and discussion sessions led any development on the participants' cognitions regarding language assessment.	reflection reports, participant reflection reports	teachers' needs regarding assessment. Almost every ten days, teachers participated in a RRD session. Before each session, teachers filled in KWL charts. They discussed the topic when they participated in the session. Finally, teachers wrote reflection reports after every RRD session.	Analysis
Post-RRD sessions- forty-one days.	Do reflective reading and discussion sessions lead to any development in EFL teachers' cognitions concerned with language assessment, as they perceive? If so, what are the changes?		Pre-semi-structured interviews, classroom observations, post semi-structured interviews.	Pre-semi structured interviews were administered to teachers. Teachers were then observed in their classroom. Finally, post-semi-structured interviews were applied to teachers. Researcher kept reflective report throughout the session for each participant separately.	Inductive Content Analysis

In the first stage, researcher aimed at finding out participants' current nature of cognitions. Firstly, semi-structured interviews were held with each participant. After the interviews, researcher observed the teachers in their actual classroom contexts. During the observations, researcher used observation forms (Appendix F) prepared beforehand based on participants' explanations in the pre-semi-structured interviews. Also, the researcher kept field notes when it was necessary. Afterwards, follow-up interviews were organized and they were specific to the participants. They were formulated considering their explanations in the

pre individual semi-structured interviews and what they did during the observation. Therefore, researcher believed that he could reveal the nature of participants' cognitions about assessment. Additionally, participants' weaknesses and strengths were tried to be discovered in the first stage of the study. To achieve this, same procedure was applied. In other words, pre-semi-structured interviews were held with each participant, then researcher observed them in their classes. At the end, follow-up semi-structured interviews were conducted. As a result, participants' strengths and weaknesses about assessment were tried to be identified. Overall, this procedure lasted twenty-eight days.

As for the second stage of the study, researcher, supervisor and participant instructors worked cooperatively in planning of the meetings. These meetings were called as reflective reading and discussion session. These sessions were held two times in a month and each of them lasted approximately between fifty minutes and an hour. Content and sequence of the RRD sessions were specified by the researcher and his thesis supervisor. Additionally, participants' ideas were also welcomed during planning the sessions. Furthermore, seven standards for teacher competence in educational assessment of students prepared by American Federation of Teachers, National Council on Measurement in Education & National Education Association (1990) were also taken into consideration while preparing both interview questions and content of RRD sessions.

To illustrate, firstly, the researcher made a literature review of assessment and language assessment. After that, he prepared first individual semi-structured interview questions. He consulted to his supervisor and some experts. Also, participant instructors read the protocols and made comment and gave feedback on them. Then the researcher made some modifications on interview questions considering the feedbacks that he received. After getting legal permission from the ethics committee of the university, each semi-structured interview was piloted systematically according to systematic approach proposed by Rubin & Rubin (2006) including editing, early pilot and full pilot. Interviewees for the pilot interviews were non-participant language teachers. After some adaptations considering wording, punctuation and sequence of the interview questions had been made, final versions of the interviews were developed and applied to the participants after approval of the supervisor. Then RRD sessions started. Before each session, participants were given some book chapters or articles which were about topic of the session. They were also provided with KWL charts and asked to fill K and W columns before reading. They were also informed about filling in L column after they had read the reading material(s). After they had been ready for the meeting, RRD session was

held. Teachers shared what they learnt with themselves. They exchanged the information and supported their learning. Meanwhile, researcher took the role of an organizer and facilitator during these meetings. At the end of each session, researcher and participants summarized what had been discussed in the meeting and the researcher made closure. Before leaving, participants were given reflective report templates and asked to reflect on. The mentioned procedure was applied separately for each RRD session. All of the sessions were audio-recorded. Detailed information about the sessions has been given in the following sessions. To put it simply, during the second stage of the study, each participant attended ten RRD sessions, filled in ten KWL charts and wrote ten reflection reports. The second phase of the study lasted a hundred and thirteen days.

In the third stage of the study, researcher tried to find out whether there was a development in participants' cognitions regarding language assessment. Two different pre semi-structured interviews were applied to each participant. They were then observed and then they were asked to attend follow-up interviews. This procedure also lasted forty-one days.

In the pre reflective reading and discussion session stage, firstly pre semi – structured interviews were held in order to reveal participants' current cognitions regarding language assessment. Their cognitions were examined with the help of such data collection instruments as semi-structured interviews, classroom observation with field notes and post semi-structured interviews. The interviews were conducted participants' own offices. Each interview was audio recorded. Meanwhile, researcher took notes while listening to the interviewees. Each interview lasted approximately twenty-five minutes and was held in Turkish in order to minimize misunderstanding. After all semi-structured interviews had been completed, each participant was observed in their classes to see whether teachers' claims and explanations match their actual classroom practices, or not. Each observation was also video recorded. Before the observation, researcher prepared observation form (Appendix F) based on participants' explanations given in the interviews. Also, during the observation, the researcher took field notes which were in line with the purposes of the classroom observation. As a last step of 1st stage of the study, follow-up interviews were held. Additionally, in the first stage of the study, participants' weaknesses and strengths regarding assessment were aimed to be explored. In line with this purpose, semi- structured interviews and classroom observations were applied. Pre-semi-structured interviews were prepared referring to seven standards for teacher competence in educational assessment of students prepared by American Federation

of Teachers, National Council on Measurement in Education & National Education Association (1990) and reviewing the literature. Then, data obtained from these sources were collected. Thus, first stage of the study was completed.

Following this, four participants were involved in the reflective reading and discussion sessions whose objective was to promote the development in participants' cognitions considering language assessment. Before starting this procedure, researcher and his thesis supervisor prepared a syllabus containing a reading list. The number of these sessions was ten. Before each RRD session, participants had been given the reading materials along with a KWL chart before actual meeting took place. Then they were asked to fill in K (What I know) and W (What I want to know) columns before reading the materials. Nearly a week was given to participants for completing their readings. After they had read, they gathered to discuss what they learnt. In this part of the meeting, the researcher took the roles of moderator and facilitator. Each session was audio recorded. After each RRD session, participants were kindly asked to fill in the L (What I have learnt) column of the KWL chart. Also, they were asked to write a reflection report. Meanwhile, a reflection report template (Appendix A) including guiding questions were given to the participants.

Ten RRD sessions were held. During these sessions, participants were introduced with basic concepts in testing and assessment in language teaching. Afterwards, they learnt types and purposes assessment. In addition, they were asked to read and discussed about traditional and alternative assessment techniques along with their implementations, strengths and drawbacks. Choosing or developing assessment techniques / methods appropriate for instructional decisions, administering, scoring and interpreting the results of both externally produced and teacher produced assessment methods were the topics of the RRD sessions. Moreover, they dealt with developing valid grading procedures, communicating assessment results to students, parents and other stakeholders, using assessment results when making decisions about students, planning teaching, developing curriculum and school improvement. Performance assessment and designing or adapting rubrics, developing subjective and objective tests: writing test items properly and ethical considerations in assessment were successively read and discussed by the participants in the RRD sessions.

As it is said before, throughout these RRD sessions, participants filled in KWL chart and kept reflection report for each session. Besides, the researcher as a moderator kept a researcher reflection report including what he learnt during the session. These sessions lasted a hundred thirteen days. After second phase of the study had been completed, third stage

started. In this phase, whether there was any development in teachers' cognitions considering language assessment was aimed to be revealed. In order to attain this ultimate aim, semi-structured interviews were applied to each of the participants. Then, teachers were observed in their classroom setting to see whether there is also a development in participants' assessment related classroom practices. Before the observation, an observation scheme was prepared for each participant based on the statements they gave in the interviews. At the same time, the researcher took field notes during observations. When all the observations were completed, a follow-up semi-structured interview was conducted with each participant once again in order to gain in-depth understanding. Actually, in the third phase of the study, two main semi-structured interview sessions were held with the participants. In the first of these interviews, the participants were asked the questions in the previous semi-structured interview which was carried out in order to determine their strengths and weaknesses related to assessment in language teaching. After the answers to these questions were recorded, the participants were once again subjected to a semi-structured interview again. In this interview, the participants were provided with questions along with their answers to these questions before reflective reading and discussion sessions. Afterwards, the participants were asked whether they wanted to add something to their answers to these questions after these RRD sessions.

All the procedures explained above implemented for all participants nearly in the same weeks. All the efforts during this procedure were for leading to a development in participants' cognitions regarding language assessment and help them put what they had learnt into their practices.

3.5.1. The Reflective Reading and Discussion Procedure

First of all, reflective reading and discussion sessions were prepared as a draft by the researcher based on the review of literature, participants' answers to interview questions, their suggestions. Moreover, seven standards for teacher competence in educational assessment of students prepared by American Federation of Teachers, National Council on Measurement in Education & National Education Association (1990) were taken into consideration in planning the content and sequence of sessions. After consulting to the supervisor and getting her approval, content and order of the sessions were decided by taking teachers' opinions into account. Sessions were conducted in the researcher's office at a time previously specified with the participants. On average, once every ten days, the participants met for reflective discussion. Each session was audio recorded after the consent of the participants was

obtained. Generally, each RRD session lasted approximately an hour. Overall objectives of the sessions were listed as below:

- a) to develop Turkish EFL instructors' cognitions regarding language assessment
- b) to detect Turkish EFL instructors' learning gains throughout these sessions.

Sessions were held in Turkish and they started in October, 2019. The syllabus of the training is presented in detail as follows:

Table 2.

Content of Reflective Reading and Discussion Sessions

Topic of the RRD Session	Date and Duration of RRD Session
Session 1: Key concepts and definitions regarding assessment, purposes of assessment, types of assessment, principles in language assessment	Session 1: 5th of October, 2019 – 32 minutes.
Session 2: Assessment techniques: Traditional ones vs Alternatives	Session 2: 19th of October, 2019 – 46 minutes.
Session 3: Choosing or developing assessment techniques / methods appropriate for instructional decisions	Session 3: 29th of October, 2019 – 49 minutes
Session 4: Administering, scoring and interpreting the results of both externally produced and teacher produced assessment methods.	Session 4: 9th of November, 2019 – 53 minutes
Session 5: Developing valid grading procedures	Session 5: 19th of November, 2019 – 56 minutes
Session 6: Communicating assessment results to students, parents and other stakeholders	Session 6: 30th of November, 2019 – 51 minutes
Session 7: Using assessment results when making decisions about students, planning teaching, developing curriculum and school improvement	Session 7: 11th December, 2019 – 54 minutes
Session 8: Performance assessment and designing or adapting rubrics	Session 8: 21st December, 2019 – 39 minutes
Session 9: Developing subjective and objective tests: writing test items properly	Session 9: 2nd January, 2020 – 42 minutes
Session 10: Ethical considerations in assessment, recognizing unethical, illegal, inappropriate assessment methods and uses of assessment information	Session 10: 12th January, 2020 – 38 minutes

Session 1: Key concepts and definitions in assessment

Within the scope of RRD session phase, participants were asked to read some reading materials about key concept along with assessment related definitions, types and purposes of assessment, principles in language assessment. Reading materials were specified according to the purpose of the session after consulting to the advisor and the participants and were distributed to the participants a week before the session. In fact, the reason why we kept the reading materials varied was to create an opportunity for the participants to choose the most suitable reading material for them. Although the number of the reading text was high, purposes of all of them was to increase the knowledge level of the participants about targetted issues. Through reading and discussion, it was aimed that participants could expand their learning trough not only reading but also supporting one another's learning through exchange of ideas.

Session 2: Assessment Techniques: Traditional vs Alternatives

Main purposes of this RRD session was to help participants have information about traditional and alternative assessment techniques. Also participants were aimed to read and discuss why alternatives came out, strengths of alternatives along with weaknesses of traditional assessment. The next aim was to explain and discuss each technique separately and show its strengths and drawbacks. Finally, we aimed to enable the participants to see the implementation of each technique and to encourage them to reflect on their own assessment practices, and as a result, to cause a development in their cognitions. For these reasons, previously determined reading texts were delivered to the participants and they were asked to read and fill in a KWL chart before the meeting. Then, this process was followed by reflective meeting in which participants shared knowledge with each other and discussed one another by reflecting on their own knowledge and actual classroom practices. This process was audio recorded and it ended with a short summary of the topics of the session. At the end, participants were asked to write a reflection report about the session and informed about the next RRD session.

Session 3: Choosing and developing assessment methods appropriate for instructional decisions

In the session three which was choosing and developing assessment methods appropriate for instructional decisions, the objective was to provide participants with

necessary information of the criteria that should be considered while choosing or designing an assessment method. Helping participants evaluate their own practices and presenting them with opportunities for making adaptations in their actual assessment practices as a result of reading the materials and discussing with each other were also aims of the session. For these purposes, KWL charts were given to the participants asked to the fill in K and W charts before reading and L chart after reading the materials. By doing so, it was aimed to improve the knowledge level of the participants on the relevant subject. In the second phase of the session, participants gathered and enriched their learning on the targetted issues through exchange of ideas and discussion. To illustrate, participants first mentioned what they had learnt as a result of reading. In addition, during the discussion session, the participants sometimes were objected to their colleague's ideas and sometimes they deepened the discussion by adding details in line with the statement put forward by other participants. With guidance of the researcher, session lasted approximately an hour. At the end, participants were thought to reinforce their learning about the issues. Then, the researcher made closure and the session ended. Finally, the participants were informed about the next session and relevant reading texts were given to them.

Session 4: Administering, scoring, interpreting the results of both externally produced and teacher produced assessment methods.

The topic of the fourth RRD session was about application of assessment methods / tools obtained from outside or produced by the teachers themselves, their scoring and interpretation of results obtained from these methods. Through this session, participants read and discussed what they need to do to safely apply assessment methods. In addition, the participants had reflective reading and discussion about what should be taken into account when scoring and how to effectively interpret the data obtained as a result of assessment practices. After readings, the participants first mentioned the weaknesses they had had or inappropriate practices they had made before the session at the beginning of the meeting. Based on what they learnt as a result of readings, the participants reported to correct their mistakes. It was observed that the participants took notes during the discussions. Throughout the session, participants exchanged detailed ideas on how to administer assessment tools and methods reliably, how to score them suitably, how to use assessment results appropriately. At the end of the discussion, researcher as a moderator of the session summarized what was discussed in the meeting and participants were asked to write reflective reports.

Session 5: Developing Valid Grading Procedures

In developing valid grading procedures session, the aim was to provide a deeper understanding of the issue of validity in both scoring and grading. Firstly, based on the statements of the participants, the focus was on definitions of scoring and grading. Also, difference between these two terms was dealt with. There was then a reading and discussion about how to score and grade validly emphasizing once again on validity. In the second phase of this RRD session, participants first explicated what they learnt from their readings. Later, they complemented each other's explanations and enriched their learning. In other words, the participants contributed to each other's learning by exchanging ideas about valid and reliable grading and scoring. In this process, researcher took the role of a guide. Sometimes, the participants asked questions during the discussion process or sometimes they asked for the researcher's approval for their explanations. Then the researcher contributed a little to the process without interfering the flow as much as possible. After the discussion meeting which lasted almost an hour, the participants were informed about the next session and were kindly asked to write a reflective report.

Session 6: Communicating Assessment Results to Students, Parents and other Stakeholders

In this session, it was first aimed to make participants aware of the importance of the assessment results. They then learnt about significant methods used to explain the assessment results, emphasizing that these results should not be announced casually. In the second part of the session, the participants talked about who are the stakeholders being interested in assessment results. At the beginning of the discussion, the participants expressed how they had announced the assessment results before reading about the topic of the session. Then, based on what they read, the participants corrected their own and their colleagues' mistakes and again touched upon how the assessment results should be shared. At the end of the session, each participant talked about what they would change in their practices after this reflective reading and discussion session.

Session 7: Using assessment results when making decisions about students, planning teaching, developing curriculum and school improvement

In this session, participant read the relevant reading materials and then they gathered for discussion. Meeting started with discussion of formative assessment. Then participants made some inferences based on their learning outcomes as a result of their reading. With this

session, the researcher aimed to improve the participants' knowledge level about the use of assessment results in order for better teaching practices. The participants were encouraged to talk about their previous practices considering the areas in which they use the assessment results. Through guiding questions, researcher aimed to figure out what they would do after the session. Participants not only mentioned the inappropriate performances or missing parts in their practices but also talked about what they would change after the session, how and for what purpose they would do it. During the discussion process, researcher observed that the participants easily exchanged ideas with each other and tried to improve their current learning. Finally, participants were kindly asked to write a reflection paper.

Session 8: Performance assessment and designing a rubric

With the performance assessment and designing rubric session, researcher aimed for the participants to have in-depth knowledge about performance assessment. Later, another aim of the session was to raise participants' awareness about importance of using rubric, how to design and use it. In order to attain these goals, participants were provided with reading materials. They then were given time to read the materials that they selected. In the second phase of the session, participants gathering for discussion and exchanging ideas stated that performance assessment was something very different from what they actually had done. They also found mistakes in their practices thanks to new relevant information they learnt. They shared their interpretations on assigned readings. They highlighted reliability and its connection to use of rubrics. During the discussion, researcher observed that the participants shared the given reading passages among themselves. In fact, the second session, started with each participants' explanation of the reading material that s/he selected to read to other participants. Afterwards, the participants talked about what they learnt, what they misunderstood and what they corrected as a result of reading. At the end of the session, the participants explained to each other what they gained as a result of the RRD session that they attended and they stated that their learning gains would be reflected in their practices. Finally, researcher as the observer and the organizer of the session, he summarized what was said throughout the session, informed about the participants for the next session and asked them to write a reflective report.

Session 9: Developing subjective and objective tests: Writing test items properly

This session was designed to help participants have knowledge of distinction between subjective and objective tests. Also, it was designed to introduce types of test items and to

inform participants about writing of these items. Thus, at the end of the session, the participants would have the competence to distinguish between subjective and objective tests, to both recognize and prepare various test items. For these purposes, after participants had read their assigned reading materials, they shared their knowledge with each other. They stated how they had prepared the exam before this session being aware of their own errors. Researcher noticed that the participant identified the errors in their practices as a result of reading. During the discussion, they exchanged information about how the test could be developed, which test items would be included in a test and how the order of items in a test would be. During the second phase of the session, the participants focused on the kinds of the item and how to write any kind of an item. They mentioned key features which should be considered while writing test items. They were also asked to reflect on their knowledge and practices on this issue. They made evaluations about not only their knowledge about item writing but also their actual item writing practices. At the end of the session, researcher assigned the participants to read a book chapter and an article about topic of the next, last, session.

Session 10: Ethical Considerations in assessment

In this last session, topic was ethics. This session targeted to help participants learn ethical considerations in every phase of assessment. They first began to focus on concept of ethics. Then they examined the concept of ethics at each stage of assessment separately. For instance, while preparing the tests, administering the tests, scoring and grading the tests, announcing exam results the participants discussed how to be ethical at all these stages. Afterwards, they gave examples of ethical, non-ethical teacher assessment practices. At the second stage of the meeting, EFL instructors reflected on their own assessment practices in terms of ethics. They honestly explained their previous practices and they also gave constructive feedback to each other by evaluating one another's practices regarding ethics. They then continued with a reflective discussion in which the participants reflect on their own practices and express their decisions for their further assessment practices.

Generally, each training session was organized and conducted based on the participants' voices. Generally, each session was divided into two in itself. In the implementation of these sessions, participants first shared the theoretical knowledge with themselves and then they reflected on their own knowledge and practices regarding the topic of the session. Throughout all these sessions, participants were encouraged to read the relevant reading materials and join

the meetings for discussion. The researcher also learnt many things and kept reflection reports including his learning outcomes as a result of these sessions. The steps in the data collection procedure of this study are given in the Appendix C chronologically.

3.6. Data Collection Tools

This qualitative study utilized different kinds of qualitative data collection tools. To ensure with – in – method triangulation, five different qualitative data collection instruments were used throughout the study. These tools which were employed before, during and after RRD sessions are listed in the following sessions.

Table 3.

Data Collection tools and their linkage to RQs

Research Questions	Data Collection Tools
What is the nature of EFL instructors' cognitions regarding language assessment?	Pre-semi-structured interviews, classroom observation with field notes, post semi structured interviews and researcher reflective report.
What are the participants' strengths and weaknesses considering language assessment as they perceive?	Pre-semi-structured interviews, classroom observation with field notes, post semi structured interviews and researcher reflection report.
Do reflective reading and discussion sessions lead to any development in EFL teachers' cognitions concerned with language assessment as they perceive? If so, what are the changes?	KWL charts, researcher reflection report, participant reflection report, pre-semi-structured interview, classroom observation, post semi-structured interview.

In the pre RRD session phase, the aims were to explore participants' cognitions regarding assessment and to figure out their strengths and weaknesses about the issue. To attain these goals, data were obtained through individual semi-structured interviews, classroom observation with field notes, follow-up individual semi-structured interviews and researcher reflection report. During the RRD session, data were collected through KWL charts, and reflective reports both from participants and researcher. In the third phase of the study aiming at to find out development in participants' cognitions, if there is any, data sources were individual semi-structured interviews, classroom observations with field notes, follow up individual semi-structured interviews and researcher reflection report.

3.6.1. Semi-structured interview

Interviews were employed as a data source throughout the study since they are the most common method of data collecting in qualitative research (Bryman, 2012). Interviews are

frequently used to obtain data on participants' thoughts, feelings, and beliefs, and they give an in-depth insight of how interviewees and interviewers interpret the world (Ary, Jabos, Sorensen & Razavieh, 2009). Furthermore, interviews allow researchers to study the phenomenon under inquiry in depth by offering opportunities for researchers to gather extra data when the responses to the original questions were not comprehended well (Mackey & Gass, 2005, p.173). According to Frankel and Wallen (2006), there are four sorts of interviews: structured, semi-structured, informal, and retrospective. Considering the objectives of the study, each interview was a semi-structured interview in which the researcher asked a series of predetermined questions related to the current specific objectives of the study. The researcher examined the literature, received support from his supervisor, and benefited from teacher competences in educational research (American Federation of Teachers, National Council on Measurement in Education & National Education Association, 1990) to design interview questions. With a preliminary list of semi-structured interview questions, the researcher consulted his supervisor once again, and the two of them agreed on the final form of the semi-structured interview procedure (see Appendices I, J, K). Semi structured interviews were employed at pre, during and after RRD sessions stages of the study. At the pre RRD sessions stage, two different semi-structured interviews were held. Firstly, researcher interviewed the participants aiming at revealing their current nature of cognitions considering language assessment then he held a semi-structured interview with each of the participants one more time in order to figure out their strengths and weaknesses concerning with the subject. As part of the post-RRD sessions phase, semi-structured interviews were undertaken to see if participants' assessment-related beliefs had changed. The researcher was very careful not to speak too much throughout the interviews in order to maximize the participants' talking time and make the most of the available time (Cresswell, 2012). All interviews were conducted in Turkish and they were transcribed verbatim manually. During the analysis of the obtained data, only extracts from the interviews which were used while presenting the findings were translated into English.

3.6.2. Classroom Observation

To collect comprehensive data which are in line with the purposes of the study, that is, to better understand current nature of the participants' cognitions about assessment before RRD sessions and to see whether any change happened or not both in participants' cognition and their assessment practices after RRD sessions, the present study benefitted from the classroom observation. As Cohen, Manion and Morrison indicate (2007), by allowing

researchers to gather information from a real-world setting, classroom observations provide researchers access to more trustworthy and authentic data. Also Yıldırım and Şimşek (2013) state that classroom observations strengthen the data gathered from participants via semi-structured interviews, allowing researchers to draw more reliable conclusions. Considering this, classroom observation with field notes was included in the data collection process of the study.

To illustrate, first stage of the study targetted to find out participants' existing cognitions considering assessment. Also, it aimed to shed light on their strengths and weaknesses regarding assessment. Bearing in mind these aims, participants were observed in their classrooms and nearly all of the observations were video-recorded. After the third phase of the study had been completed, participants were observed once again in order to detect a development in their practices in terms of assessment or not. In other words, aims of the observations at the last stage evolved around finding out whether participants apply what they claimed to practice in their actual assessment practices. During these observations, the researcher took a non-participant role. Based on each participants' explanations in his/her interviews, an observation scheme was prepared for observing a specific instructor by clearly giving information about physical environment in which observation would take place as Yıldırım and Şimşek (2013) suggest. As a result, observation schemes were different from one another. Researcher utilized them during the observation and he also kept field notes which were in line with the aims of the observation by keeping them as concise, descriptive and detailed as possible. Each classroom observations lasted almost an hour. Finally, researcher himself wrote his reflection in the researcher reflection report.

The objective of the classroom observation in this study was not to collect quantitative data but to gain a comprehensive understanding of the investigated phenomena. Qualitative data were obtained from the classroom observation. The data obtained were subjected to the content analysis.

3.6.3. KWL Chart

A KWL chart is a type of visual organizer that can help readers express what they know before reading, what they want to learn through target reading, and what they learn after reading in an arranged and systematic manner.

In addition to its usage as a reading strategy, the KWL chart is an effective tool for promoting cognitive activity before, during, and after reading (Ogle, 1986). In this study,

KWL chart was utilized in order to guide participants throughout their reading process. In addition, KWL charts were used to systematically obtain relevant data about the participants' prior knowledge of the subject before reading and the improvement in their knowledge level after reading, if any. During the data collection period, participants were given a KWL chart prior to each RRD session. Next, they were asked to fill in K column standing for "what I know" before reading the materials. By doing this, it was aimed to reveal participants' pre-knowledge about relevant subject before each session. Also, participants had been asked to fill in W column which stands for "what I want to learn" before they started reading the assigned materials. The purpose of completing W column was to make participants aware of the areas that need improvement both in their knowledge levels and in their practices. Finally, the participant instructors were asked to fill in L column after reading. L column stands for "what I have learnt". In line with this explanation, the participants were asked to talk about contribution of reading to them in column L, that is their learning outcomes as a result of the reading. In column L, participants were asked to write not what they understood from the post reading materials, but rather what they learnt for this time during the reading process.

The KWL chart is recognized for its efficiency in communicating the knowledge gained through reading, indicating the unknown, and revealing what should be learned (Johnson, Johnson & Ray, 1984). KWL charts, according to Car & Ogle (1987), offer an opportunity to display the knowledge gained via reading and prior knowledge related the topic of reading in a thorough and logically ordered way. Furthermore, Ogle (1986) asserted that the KWL chart allows users to take an active role in the reading and learning processes mentally. Considering these, KWL chart was included and effectively utilised throughout the second stage, during RRD sessions, of the study.

3.6.4. Reflection Reports

Another source of data is reflection reports written by both participants and the researcher himself. According to Cengiz and Karataş (2014), reflective writing provides individuals with the opportunity to self-evaluate by examining the subject in various and different ways. Additionally, Dewey (1993) indicated that researchers can benefit from reflective essays in order to obtain information about participants' thoughts, interpretations on a certain phenomenon. Based on these explanations, for the current study, participants were asked to write a reflection report about each RRD session. Some guiding questions about perception of any RRD session, its effective and ineffective sides, what participants learnt during the session, how they could apply what they learnt, their learning outcomes, in the

session into their actual assessment practices were given to the participants. These questions were believed to help participants while filling in their reflection reports. Then obtained data were analyzed based on content analysis and relevant research questions were answered.

3.7. Data Analysis

The present study firstly aimed to reveal participants' nature of existing cognitions considering language assessment. Then it also examined the participants' strengths and weaknesses regarding language assessment. Based on participants' weaknesses, as they expressed, organizing RRD sessions was one of the aims of the current study. Moreover, the study aimed at developing participants' cognitions through reflective reading and discussion. Finally, this dissertation sought to explain whether there was any development in EFL instructors' cognitions and practices of language assessment, or not. In order to accomplish these aims, the current study took a qualitative stance toward data collection and analysis. Considering qualitative nature of the study and its purposes, data collection tools, content analysis was employed in order to analyze the collected data. During the analysis, the frequency of citation of each emerging code was considered.

This part will introduce the analysis process and procedure followed in generating and discovering codes, sub-themes, and themes from the raw data. As Cresswell (2009) indicates, the process of data analysis entails preparing the data for analysis, performing analysis to make sense of the text, and interpreting meaning. Similarly, Tavakoli (2012) describes the data analysis as a "process of reducing accumulated data collected in research to manageable size, developing summaries, looking for patterns, and performing statistical analysis" (p.145). Based on these views, it might be asserted that the primary aim of a qualitative researcher is to investigate a case or an event in the data rather than making broad predictions about the entire entity.

Instead of presenting gathered data with numerical explanations, qualitative data analysis encompasses research methodologies that try to disclose the underlying hidden meaning through non-numerical analysis and interpretations of observations. (Babbie, 2007). Qualitative researcher is responsible for "interpreting a case or situation rather than making predictions about it" (Bhattacharjee, 2012, p.113). In this study, it is thought that this individual meaning extraction process could be carried out through content analysis.

Krippendorff (2004, p.18) defines content analysis as "a research technique for making replicable and valid inferences from texts to the context of their use". Also, Weber (1990)

indicates that “content analysis is a research method that uses a set of procedures to make valid inferences from the text (p.117). It is a method of coding, categorizing, comparing and contrasting, and making inferences from a text (Ezzy, 2002). Additionally, content analysis is defined by Cohen, Manion and Morison (2018) as a “process of summarizing and reporting written data – the main contents of data and their messages (p.674). It seeks to scientifically and methodically understand a reality, as opposed to just counting the words in a targeted text. (Zang & Wildermuth, 2009).

In light of the definitions provided, it is possible to conclude that content analysis is a method that allows valid interpretations and conclusions to be drawn from the targeted text in a systematic and scientific manner. According to Krippendorff (2004), content analysis gives the researcher with new insights that assist him or her better understand the case, phenomenon, or event under investigation by informing his or her practices. As a consequence of a set of reflective reading and discussion sessions, it can be stated that content analysis leads researchers acquire critical understandings regarding the evolution of instructors' assessment-related cognitions.

While analyzing the qualitative data derived from semi-structured interviews, KWL charts, non-participants classroom observations with field notes, reflective reports, the steps below suggested by Creswell (2013) were taken by the researcher into consideration which are as follows:

Step 1: organizing the data

Step 2: reading and memoing

Step 3: describing, classifying and interpreting the data into codes, categories and themes

Step 4: interpreting the data

Step 5: representing and visualizing the data (p.182-188).

As stated before, this current study includes three phases such as before RRD session, while RRD session and after RRD session. In the before RRD session, first two research questions were generated. These research questions were formulated in order to shed light on participants' current nature of cognitions regarding assessment. Also their strengths and weaknesses were aimed to be discovered considering assessment. In order to attain these goals and find answers to the first two research questions, semi- structured interviews, non-participant classroom observations, follow-up semi structured interviews, researcher reflective

reports were utilized as sources of data and they were analyzed via content analysis using Maxqda from an inductive point of view. Saldana (2009) mentions several advantages of using a software while analyzing the qualitative data by saying that “software efficiently stores, organizes, manages, and reconfigures your data to enable human analytic reflection” (p.22). Also, the software facilitates visual organization of the analysis. Software is advantageous for a researcher in terms of practicality regarding showing the key words or phrases and similarly coded data for the investigation. Moreover, Saldana (2009) explicates that software provides researcher with an opportunity to move back and forth between multiple tasks systematically. Bearing these advantages and intensity of the qualitative data in the study in mind, Maxqda 2020 was utilized as a software for the analysis of data of the current study.

When comes to while RRD session, research reflection reports, KWL charts, participant reflection reports were used as data collection instruments and the data derived from these tools were analyzed based on content analysis. These investigations were carried out to answer 3rd research question which aimed to unravel whether there is any development in the nature of participants’ cognitions about assessment. As for the after RRD session, the third research question that was formulated was also relevant for this phase of the research. In order to answer it, corresponding tools were semi-structured interviews, non-participant classroom observation with field notes, follow-up semi- structured interviews. As in two previous phases, all the obtained qualitative data were analyzed based on content analysis by using a software called Maxqda. While reporting the data analysis, researcher demonstrated total results in order to provide with a broader representation of the phenomenon under scrutiny.

The following sections provide with a detailed explanation about content analysis process.

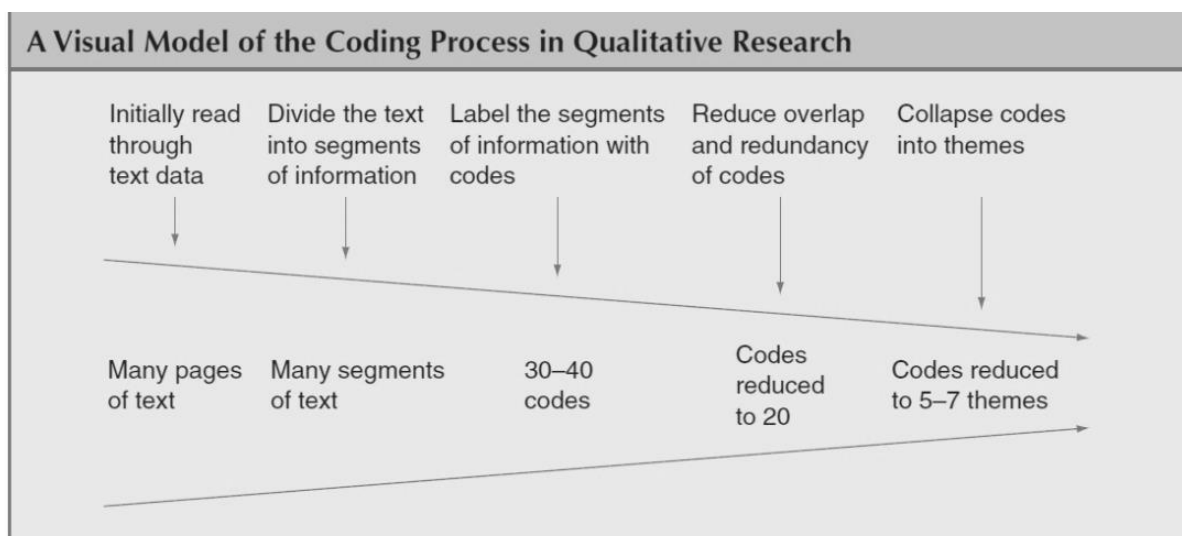


Figure 2. The steps taken in the analysis of the qualitative data (Creswell, 2012, p.244)

This figure also explains the overall procedure of application of content analysis of the data collected for the current study.

3.7.1. Transcription

Transcription is an essential tool for a researcher to become intimately familiar with a text. (Dörnyei, 2011). Researchers benefit from transcription in multiple ways, including to show obtained data more systematically and to use the data for new investigations or procedures (Bryman, 2012). In order to analyze the spoken data obtained via semi-structured interviews as accurate as possible, all the interviews were transcribed verbatim. The process of transcribing verbal data is thought to be more efficient than some other options like constantly listening to the texts from the recorder to identify and note key points (Gillham, 2000). This method is an important first stage in qualitative data analysis because it enables the researcher to concentrate on certain aspects of a text, even though it sometimes may lead to confusion and cause the researcher to lose sight of the relevant parts of the data. (Green, Franquiz & Dixon, 1997).

In the current study, recordings obtained through semi-structured interviews during before RRD session and after RRD session stages were transcribed separately. After all, the transcripts had been completed, the researcher read all of them to check the texts considering spelling, punctuation and grammar. Although transcribing is viewed as time-consuming and difficult, it is an important approach for a researcher to become familiar with the material (Riessman, 1993). As a result of these processes, researcher not only prepared the data for the

analysis but also started to familiarize himself with the whole data set as this process called as pre-coding by Dörnyei (2011).

3.7.2. Coding

First of all, the qualitative data was prepared for the analysis in separate folders. Data derived from semi-structured interviews, KWL charts, reflection reports, and researcher reflection reports were read many times successively. Therefore, researcher immersed himself into data, familiarized with them and had a holistic view (Creswell, 2003). While reading the transcripts, researcher wrote down some notes in the form of memos. This note-taking helped the researcher create new ideas based on his initial understanding and interpretation. The researcher re-read the whole data set several times, which resulted in potential revisions in memos, which helped the researcher develop codes for semi-structured interviews, reflection reports written by both participants and the researcher, KWL charts, and field notes.

Coding is the process of organizing text and highlighting key elements of data by labelling them (Cabaroglu, 1999). Also, coding is described as a significant component of a qualitative data analysis referring to “a process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes” (Kothari, 2004, p.123). Coding is the process of transforming raw data into a standardized format that may be utilized for computer-based analysis. (Babbie, 2007).

Keeping in mind that since qualitative data analysis depends on the interpreter and the context of the study, it cannot be governed by universal standards and may represent differences as well as changes as the inquiry progresses and the data develop. (Coffey and Atkinson, 1996). Furthermore, it should be noted that the texts convey subjective interpretations, implying that they do not provide objective reader-independent conclusions; rather, they present a diversity of interpretations. (Krippendorff, 2004). Codes were constructed according to three criteria: 1) a word / sentence that is repeatedly uttered by the interviewee 2) a word / sentence that seems interesting or new for the study 3) a word / sentence that is already found and common in the literature. Considering these, reoccurring words were carefully recognized and categorized into codes and categories throughout the texts. (Bhattacharjee, 2012). Codes are generated in two ways. Firstly, researcher codes the meaningful segments based on a predetermined template of codes. Thanks to this predetermined template, codes are specified deductively. Also, codes which do not exist but are formulated and elicited from a text by using inductive reasoning in a process which is called inductive coding (Patton, 2002; Babbie, 2007; Bryman, 2012). For the current study,

the latter approach was employed. It means that the researcher initiated with particular data units and then created codes, categories and themes to constitute a broad picture of the case under examination.

Upon receiving all the data set, the researcher read the entire data many times in order to have an idea about the data. Saldana (2009) claims that this process is helpful for researchers in terms of reflecting on the content of the data and having sense of belonging for the data. Therefore, researcher would have a deeper understanding about the data, have an idea what is going on in the data which makes it easier for the researcher to analyze and evaluate the data. (Thornberger & Charmaz, 2011). After the researcher had become acquainted with the data sufficiently, he divided text into segments of information which are called as text segments (Creswell, 2012). At this point, the researcher reached out to a large number of segments and labelled them with appropriate codes so that he could have more structured units of analysis for data reduction purposes. Meanwhile, it should be pointed out that some parts included in the texts that were irrelevant to the aims of the study and research questions were ignored throughout the analysis. Following that, the researcher occasionally gave codes to the significant units by utilizing words or phrases expressed by the participants, a process known as in-vivo coding (Saldana & Omasta, 2018). Codes are defined by Miles and Huberman (1994) as “tags or labels for assigning units of meaning to the descriptive or inferential information compiled during during a study” (p.56). As a result of first cycle of coding, a number of codes were generated and Corbin and Strauss (2008) define this process as open coding. Later on, these emerging original codes were counted, examined in detail and reorganized by clustering them with other codes to form collective categories, which were then rearranged in order to be grouped under unified themes within the framework of the study. At the end of these processes, researcher had a clear idea of what themes, categories and codes along with extracts are and how they connect to one another. While reporting the analysis of the study, each theme was explained along with categories included in it. Excerpts taken that is direct quotes from the interviews were provided with by translating them into English. Gall et al. (2003) expressed that “direct quotes of the remarks by the case study participants were particularly effective because they clarify the emic perspective, that is, the meaning of the phenomenon from the point of view of the participants” (p.469). Quotes make it simple for readers to follow the conclusions of study. In addition, a chronological technique was used to summarize the findings of study findings (Yin, 2003) in order to assist readers in interpreting the proposed premises step by step throughout the entire procedure. The analysis

of emerging codes, categories and themes were based on constant comparison method defined by Maykut and Marehouse (1994) as a method of qualitative data analysis in which all relevant units produced from the acquired data are compared to one another iteratively. In other words, all the meaningful segments emerging from the data are coded. Then they are compared to all other codes. Afterwards, based on the relations among codes, these initial codes are categorized, grouped, and continually enhanced (Maykut & Marehouse, 1994, p.347). The basic goal here is to restructure the collected data and generate "a reasonable reconstruction of the data" (Lincoln & Guba, 1985, p.347). As a consequence, the researcher would have knowledge about the investigated situation, event, or individuals in order to interpret the data. In order to achieve this aim in this current study, during the data analysis process, researcher read the entire data set many times, took some notes and coded the data by keeping memos. Then, emerging codes were continuously compared to other codes by moving back and forth in the data until no new patterns emerged and they classified into as categories. Finally, the generated categories were compared among themselves by taking into account their similarities and differences and as a result, themes were created which were in line with the research questions. During the analysis process, special attention was paid to the interrelationships and connections among each data source in order to avoid the risk of treating the data sources as separate entities when reporting the outcomes of the study. The reasoning for this is that case studies need the integration of several data sources in order to completely comprehend the whole case, rather than simply the individual aspects of the case. (Baxter & Jake, 2008).

Since the present study was mainly a qualitative study in which qualitative coding techniques were employed, it is crucial to involve multiple coders into coding process to ensure the trustworthiness of the study. As for increasing the trustworthiness of the data analysis process, the researcher applied intra-coder and inter-coder reliability. Twenty five percent of the collected qualitative data (semi-structured interviews from each stage of the study, KWL charts, reflection report) were coded by two researchers one of which from the field of English Language Teaching and the other one is from the Computer and Instructional Technologies Education. Both of them are PhD candidates in their own fields and have qualitative data analysis background. In order to calculate the inter coder reliability, the formula proposed by Miles and Huberman (1994)'s was utilized. In other words, the number of agreements was divided into the sum of agreements and disagreements in coding. As a result of the analyses, 80% consistency was found and it might be interpreted that inter-coder

reliability that is the consistency among coders is high (Saldana, 2009). Also, it should be noted that member-checking technique was applied during the data analysis process in order to ensure true conformity with the data submitted by the participants as it is claimed by the Lincoln and Guba (1985) that this technique increases the validity of the data collection tool and it also increases the trustworthiness of the findings emerging from the study (Creswell, 2012). In this study, researcher came into contact with each of the participants and asked him / her decide whether the interpretations made were reflecting their own beliefs or not. In addition to these, in order to increase the validity, data sources utilized in the current study were triangulated. As Mackey and Gass (2005) advocate, “using the technique of triangulation can aid in credibility, transferability, confirmability and dependability” (p.181). As a result, member checking showed that excerpts and interpretations made provided sound evidence.

3.8. Ethical Considerations

Research ethics is an important consideration while designing and carrying out a study. Various types of problems resulting from the nature of the study, setting of the research, procedures of the research, the techniques of data collection and the nature of the data obtained, what is to be done with the data, and participants might generate ethical concerns. In order to conduct research on people, a researcher must uphold certain ethical obligations throughout the study. Particularly qualitative researchers contact with people, thus they must be aware of and follow ethical guidelines. To illustrate, since this study is a qualitative case study, researcher sometimes has delved into participants’ lives. As a result, researchers must protect the rights, privacy, and well-being of the individuals and communities under study (Lune & Berg, 2018). Although it is a frequent misunderstanding that researchers should address ethical problems during data collection, ethics should be considered at every stage of the research process (Cresswell & Path, 2018). This may be accomplished by ensuring that research does not hurt participants, which can be accomplished by assuring data security, objectivity, informed consent, anonymity, and confidentiality. (Borg, 2001). These principles were adopted in order to avoid ethical concerns. These principles have been explicated below.

3.8.1. Informed Consent

Diener and Crandal (1978) defined informed consent as the processes in which participants have the choice to engage in the study or not after being informed of its details. According to Yıldırım and Şimşek (2018), informed consent is the voluntary engagement of

participants in a study as a consequence of growing understanding about their responsibilities throughout the study. As such, an informed consent is a prerequisite for a study in terms of voluntary participation. In order to achieve voluntarily participation credible, meaningful, detailed explanations should be given to participants. Also, informed consent form should include information about brief explanation of the nature of the research, purpose and procedure of the research, what topic will be discussed, what actions will participants be expected to perform. In addition to these, who will view, read or hear the study, possible risks and how these risks will be analyzed and prevented, how these results will be reported, where the results will be used should be explicated in detail. In addition, participants should be assured that their participation is voluntary and that they are free to withdraw from the research at any time without justification. Here, the main purpose is stressing that consent to participate in the study is based on voluntary decision as a result of being detailed informed. Additionally, practices for ensuring anonymity and confidentiality should be included in the informed consent form.

To address issues aforementioned and in order not to lead any ethical failures to occur, researcher behaved carefully and worked out all the details of research design in advance. Firstly, researcher first had obtained the approval of ethics committee of the university (See Appendices A, H) before he started to conduct the study since Ethics committees are critical for determining the soundness of research ideas. (Uwe & Flick, 2009). Prior to implementation of the study, instructors were given detailed information about the research and they were asked to participate to the study voluntarily. Then the researcher prepared an informed consent form (Appendix G) including aims of the study, procedure of the study, what kind of data collection tools will be employed, how the data will be collected and analysed, researcher's role, participants' role, where and how the results will be used, etc. Also, researcher explained that participants could resign themselves from the study whenever they wish. After checking the consent form, researcher distributed the consent forms and asked participants to read and sign it (Appendix X). As a result of these processes, six instructors accepted to take part in the study. During the data collection process, two of the participants withdrew from the study. Study was completed with four participant instructors.

3.8.2. Anonymity and Confidentiality

Participants should be persuaded by the researcher that they won't experience any harms by participating in the study. Harm, in this sense, can be physical, emotional and psychological (Berg, 2008). In order to abstain from this risk, researcher should enhance

confidentiality and anonymity. Although these two constructs are sometimes used mistakenly as synonyms, they have distinct meanings (Lune & Berg, 2018). Confidentiality means making a conscious effort to eliminate anything from the research records that could reveal who the participants are.

According to Uwe & Flick (2009), confidentiality also implies that information about participants can be utilized in such a way that no one can identify the participants based on the information provided. In addition to these, Yıldırım and Şimşek (2018) indicate that physical context in which the study is conducted and any information that may be used for identification should not be directly or indirectly deciphered. Anonymity is slightly different from the confidentiality. The essence of anonymity is that information provided by participants should in no way reveal their identity. Participants should not be identified from the information or explanation given to the research. In most qualitative research, however, since participants are known by the researchers, it might be difficult to fully ensure anonymity because a participant agreeing to face to face interview can in no way expect anonymity. At most, interviewer should promise confidentiality.

Confidentiality was maintained in this research. Firstly, the researcher prepared a detailed informed consent form and guaranteed that data will be protected and participants' identity will not be identified throughout the study. Interviews were conducted in the researcher's office. Also, some precautions were taken in case of interruption. Researcher and interviewee came to the agreement before start of each interview that in case of an interruption recording will be paused and then was resumed following interruption. Before transcription of interviews, the researcher prepared a list including nicknames of each participant in order to create consistency and the researcher systematically adhered to this list while transcribing the records and writing up the results. Therefore, participants' real identities were not revealed. Keeping in mind that giving nicknames or pseudonyms cannot completely minimize the danger of identification, the researcher took care not to provide any information about the participants' identities.

3.8.3. Data Security

Protecting data is also a significant issue that a researcher should take into consideration in order to ensure that information does not accidentally fall into the wrong hands or become public. Holmes (2007) provides some important and useful suggestions about how to protect participants' data. Her tips include not storing participants' names or any information which can be helpful for a foreigner to identify participants' identities on hard drives guaranteeing

that transcriptions do not include participants' names. Also, keeping the records of interviews namely transcripts in a blocked folder in order to achieve data security is one of the suggestions. First of all, in order to enhance data security, the researcher started with careful research design. In other words, during the stage of the study, researcher safely considered what actions will be taken to safeguard the identities of the participants as well as the data once they are collected, used, stored. Then, researcher stored each record in a folder which were kept in a separate external hard disc. As for the transcriptions, researcher kept them in a locked cabinet, when they were required to use as hard copies. Therefore, not only records but also transcripts were not unlocked by any person other than the researcher. In other words, the researcher kept whole data in a safe and completely secure container so that no one could access the data.

3.8.4. Honesty and Trust

Another principle which must be obeyed is being honest to the target participants about every details of the study. Therefore, researcher can gain trust of the participants and they can participate to the study voluntarily. In order to achieve this, researcher firstly took the official permission to undertake the study in the context of the study. Since the research will extend over months and the research procedure will be a bit complicated, the researcher prepared himself well in order to gain access to research site and gain participants' trust. In order to attain these goals, researcher first clarified his own mind about the nature and scope of the research. In other words, the researcher started with introducing himself to the participant instructors and explained his roles that he will take throughout the study. Then, he introduced aims of the study, research questions, data collection tools, how the study will be conducted, what kind of data will be collected and how they will be analyzed and used. The researcher had explained the procedure in a detailed way before the actual study started. Also, the researcher honestly explained what was expected from the participants. He tried to make participants have a total picture of what the study will entail. Additionally, he guaranteed that the procedure would not harm participants by explaining any potential risks and benefits from their participation. According to Borg (2008), one of the fundamental pillars of ethical social scientific research is the concept of do no harm which literally refers to avoiding physical and emotional or psychological harm. Keeping this in mind, researcher tried to minimize disturbance both to participants themselves and to participants' relationships with the environment. All explanations were made in Turkish language in order to avoid any

misunderstanding. Last but not least, researcher informed the participants to withdraw from the study whenever they want without providing with any justification.

3.9. Trustworthiness of the Study

Every qualitative study carried out needs to provide credible and accurate results. Increasing the credibility of the results obtained in qualitative studies requires meeting a set of necessary specific tactics used to test and verify the quality of the results obtained from the study (Miles et al. 2014). Trustworthiness in qualitative studies is defined by Given (2008) as follows:

In qualitative research, trustworthiness has become an important concept because it allows researchers to describe the virtues of qualitative terms outside of the parameters that are typically applied in quantitative research. Hence, the concepts of generalizability, internal validity, reliability, and objectivity are reconsidered in qualitative terms (p. 896).

Accordingly, Patton (2015) claimed that trustworthiness is significant for both quantitative and qualitative studies. Also he added that “criteria specifying the trustworthiness differ between two research types such as credibility, transferability, dependability, and confirmability are preferred for qualitative studies while internal validity, external validity, reliability and objectivity are preferred for quantitative studies” (p.815). Also, Lincoln and Guba (1985) mentions credibility (alternative to internal validity), transferability (alternative to generalizability), dependability (alternative to reliability), and confirmability (alternative to objectivity) pointing out that these four construct increase the trustworthiness in qualitative studies. These four criteria were employed in the current study to ensure that findings were credible and data analysis process was accurate. They are explained in the following paragraphs.

Credibility: According to Lincoln and Guba (1985), credibility refers to the extent to which the research setting is reflected in the research findings and the extent to which the researcher discovers what he or she set out to uncover in the study. Also, Patton (2015) defines the credibility as a “criterion which is concerned with whether the research findings are credible to the target population examined by the research and whether the final report of the research reflects an authentic portrait of the subject matter” (p.815). In the light of the definitions made, the credibility of the study was tried to be increased by using following strategies put forward by Lincoln & Guba (1985). Firstly, the researcher was familiar with the

culture of the organization in which the study would be carried out because he had been working there almost for a year. Secondly, data collection sources were triangulated. In other words, evidences from different individuals, different types of data (interviews, researcher reflection report, participant reflection report, KWL charts, non-participant classroom observation with field notes) and different methods of data collection (reports, observations, and interviewing) were corroborated in order to be sure that an accurate understanding of the research topic was obtained. Then, each of EFL instructors participating to the study was guaranteed that they could leave from participation in order that only those who were eager to provide genuine data would be involved. Also, participants signed a consent form for the participation to the current study and data security together with protecting participants' privacy was ensured. Additionally, peer scrutiny which is advised for researchers since it prevents the researcher from inhibiting his ability to view the research (Lincoln & Guba, 1985) was utilized in designing the research, data collection process and analyzing the data. Next, as Creswell and Miller (2000) indicate, researcher involved in member checking process in which researcher took the findings of the study back to the participants and consulted to them whether interpretations made were realistic and representative of their ideas. Therefore, participants verified the inferences made in the data analysis. What is more, the researcher provided thick description of the findings which in turn helped both himself in terms of reflection on the data collected and readers understand the phenomenon under investigation better.

In addition to the strategies (Lincoln & Guba, 1985) mentioned above, details of the data collection along with data analysis process and the role of the researcher were provided to increase credibility. While talking about the findings of the study, necessary quotations from the data were used to clarify the findings. Finally, during the whole study, the researcher debriefed continuously with his supervisor who have studies in the field of qualitative research. Thus, the researcher had the opportunity to discuss the strengths and weaknesses of the research at every stage of the study. It is thought that benefitting from all those steps, this current study was as credible as possible.

Transferability: The trustworthiness of the study is also determined by its transferability in different contexts. Transferability is concerned with the generalizability of the qualitative study's findings and interpretations to bigger groups or various situations (Bryman, 2012). However, it should be expressed that qualitative research focuses on understanding specific phenomena in their own contexts rather than generalizing the findings to the universe.

Consequently, it is frequently challenging to apply the study's findings and conclusions to bigger groups or various situations (Shenton, 2004). To improve transferability in qualitative investigations, a detailed description of the research setting, people included in the study is provided, which may provide an appropriate comparison with other contexts and samples. (Lincoln & Guba, 1985; Miles et al., 2014). In this study, researcher tried to provide sufficient, descriptive details about phenomenon under inquiry, participants, research context, RRD sessions, data collection instruments as well as data analysis. All the steps taken throughout the research were elaborated in detail so that other researchers could make adequate comparisons and decide on the transferability of the study.

Dependability: Dependability refers to how findings of a qualitative study are logically consistent with the data obtained (Lincoln & Guba, 1985). To fulfill the dependability criterion, researchers need to provide a thoroughly elaborated explanation of research context and the relation among the participants. Additionally, Miles et al., 2014) offers some points for ensuring dependability as follows:

- Clear and well-defined research questions that are congruent with the features of the research design,
- An explicit amount of the researcher's role and stance throughout the research,
- Well-established parallelism across the elements of research design (participants, settings, dates),
- Documentation of peer or colleague review (p.272).

In the current study, in order to meet dependability criterion, firstly, relation between research questions and data collection tools and the rationale behind their selection was presented in methodology chapter. Secondly, researcher role was clearly explained considering every step of the research. Next, how the data were collected along with implementation of RRD sessions session by session and data analysis were explained in detail. Furthermore, participants were informed about the analysis of the data and were asked to review the codes, categories and themes elicited from the data after data had analyzed by the researcher. Moreover, researcher enhanced triangulation of the data and utilized member checking so as to increase dependability of the data (Pitney, 2004). Last but not least, the researcher used a software program, Maxqda 2020, in order to create a data base while organizing and analyzing the data considering that it is significant to generate a case study

database in order to organize the recordings in terms of their sources, date of their collection, storage (Wickham & Woods, 2005).

Confirmability: Confirmability is important in qualitative investigations since it refers to the fact that the conclusions of the study are the result of the participants' experiences rather than the researcher's choices (Shenton, 2004). Also, Lincoln & Guba (1985) explains that confirmability indicates how much a qualitative study's conclusions are based on the experiences and views of the individuals under investigation, rather than the researcher's preferences. Additionally, Miles et al., (2014) defines confirmability in qualitative research as “neutrality and reasonable freedom from unacknowledged researcher biases” (p.272). Considering these, objectivity might be interpreted as a prerequisite for a qualitative study especially in data collection along with their analysis and reporting the findings. During interviews, researcher should be cautious while listening to interviewees and comprehending their responses to interview questions in order to decide on whether the answers given to the questions are sufficient or not. Furthermore, interviewers should avoid attempting to show their mastery over the interviewees by probing the questions in a neutral manner in order to obtain objective responses from the participants rather than imposed responses (Kubanyiova, 2015). So as to fulfil the confirmability criterion, in the current study, researcher took a number of precautions during each of three phases of the study. First of all, triangulation was employed so as to promote the confirmability in the present study. Also, intrusion of researcher bias and subjective predispositions into data collection and analysis procedures were tried to be prevented. It was guaranteed that results of the study are in line with the participants' opinions and explanations rather than researcher's own preferences. To prove this, necessary excerpts from interviews, reflection reports, KWL charts and researcher reflection reports were directly presented while explaining critical points. Moreover, participants were given the freedom of choosing the medium of interviews, reflection reports and KWL charts (i.e. Turkish vs English) throughout data collection process. During the interviews, participants chose their native language, Turkish, as the medium of the interviews. However, it was observed that participants expressed themselves in English while filling in the KWL charts and writing reflection reports. Throughout the interviews, the researcher paid great attention while listening to interviewees and added some follow up questions in order to better understand the responses given. While asking the questions, researcher was careful for using a neutral tone of voice in order not to impose his proficiency on the participants.

3.10. Researcher Role

In this study, researcher took various roles. As the study was a qualitative case study, researcher had crucial responsibility from data collection to analysis and writing up the results. That is, he conducted all the semi-structured interviews and classroom observations by himself, joined all reflective reading and discussion sessions. At the beginning of the data collection process, pre-semi-structured interviews were conducted with each of the participant. According to Yıldırım and Şimşek (2018), researchers in qualitative studies should make decisions at every stage of the investigation since qualitative studies do not adhere to any pre-determined standard procedures or specific guidelines that can be applied to varied populations and circumstances. Therefore, before the interviews researcher took role of an organizer and decision maker considering place and time of the interviews to be held and how these interviews would be scheduled. During the planning, he did not have difficulty because not only participants but also researcher himself was teaching at the research site.

Second, Cresswell (2012) states that qualitative studies are interpretive studies demanding from the researcher to be in contact with the participants continuously. Yıldırım and Şimşek (2018) note that qualitative researchers employ their social skills, interacting with individuals and using different communication tactics dependent on the setting. For this reason, the researcher negotiated with all the participants throughout the study. For instance, they worked collaboratively in terms of determining place and time of interviews and reflective reading and discussion sessions. Moreover, the researcher took an active listener role throughout the interviews by listening to participants attentively. At the same time, he constantly analyzed the data obtained from the participants, took notes and asked some follow-up questions to have further explanations from the participants when something stated was unclear or inappropriate for the interview questions. It is due to the fact that, the researcher took the role of facilitator in the interviews. As for the classroom observations, researcher took a role of a non-participant observer. In other words, he did not interfere in the flow of the lesson and took field notes when they were necessary. In addition to these, researcher again took the role of a facilitator throughout the reflective reading and discussion sessions. He moderated the sessions, he asked some further questions to help participants discuss the issues deeply, gave some minor clues helping participants remember what they read. He encouraged instructions without behaving like a prime knower of the issue of each session. At the end of each session, the researcher was on the stage so as to wrap up the session.

As for the analysis of the obtained data, firstly the researcher took the role of a transcriber. He transcribed all the interview records verbatim. Then he immersed himself to meaning making process of the study as Şimşek & Yıldırım (2018) explicate.

In general, as Saldana and Omasta (2018) state, researcher adopted a peripheral role. That is, he did not directly attend to the discussions and influenced the participants. Most of the time, researcher was silent and sat away from the participants and took an active role only when he was provided with a question by the participants.



CHAPTER IV

FINDINGS

4.1. Introduction

This chapter presents the results of the study. It had four foci. The first focus was to discover EFL instructor's nature of current cognitions concerning language assessment. Secondly, their perceived strengths and weaknesses considering language assessment were tried to be explored. The third focus was to conduct reflective reading and discussion sessions through which teachers share their ideas. Lastly, the study aimed to identify the self-perceived development of instructors' cognitions about language assessment as a result of RRD sessions, if any.

The first section of this chapter reports on the instructors' current perceptions of assessment before reflective reading and discussion sessions. It provides results from pre-semi-structured interviews, classroom observations with filed notes, post-semi-structured interviews and researcher reflection reports.

The second part of the chapter documents the results of the analyses of semi-structured interviews, classroom observation with field notes, post-semi-structured interviews and researcher reflection reports. Finally, the third part of this chapter provides data analysis results from two separate pre-semi-structured interviews, classroom observations with field notes, KWL charts, and researcher and participant reflective reports. It aims to present self-perceived development in instructors' cognitions considering language assessment as a result of participating in RRD sessions.

In general, while presenting the findings, all codes belonging to each sub-theme are included in the tables. However, in the explanations below the tables, the codes having the highest and lowest frequency of citation and being perceived as important to mention by the researcher have been shown with their quotations. In the light of these explanations, this current chapter demonstrates analyses of obtained data for each research question separately. The study's findings will be shown in the order of research questions.

4.2. What is EFL Instructors' Nature of Cognitions Regarding Language Assessment?

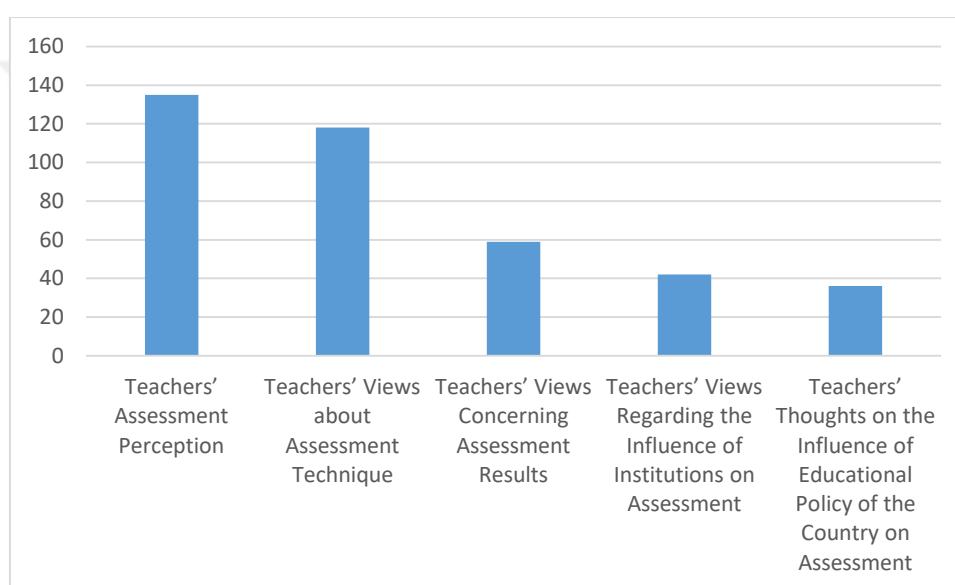
The first research question aims to reveal EFL instructors' current nature of cognition about language assessment before reflective reading and discussion sessions. To investigate their current cognitions, participants were asked to participate in semi-structured interviews during which they felt free to answer the questions. Later participants were observed during their lesson. Meanwhile, the researcher kept field-notes. Finally, post-semi-structured interviews were applied to the teachers. The obtained data were transcribed verbatim and subjected to inductive content analysis (Cresswell, 2012). Table 4 demonstrates five themes elicited as a result of the elaborative coding process.

Table 4.

Themes and Sub-themes Referring to Participants' Current Nature of Cognitions Regarding Assessment

Theme	Sub-themes
Teachers' assessment perception	<ul style="list-style-type: none"> - Teachers' definition of assessment - Teachers' views on classroom assessment - Teachers' purpose of assessment - Teachers' views of ideal time for assessment - Teachers' views on effective assessment - Teachers' initial perception of assessment and change in it over time
Teachers' views about assessment technique	<ul style="list-style-type: none"> - Teachers' assessment technique knowledge and preference - Teachers' purpose of using assessment technique - Reason(s) behind teachers' preference for assessment technique - Teachers' perception of effective assessment technique and its characteristics - Teachers' perception of alternative assessment
Teachers' views concerning assessment results	<ul style="list-style-type: none"> - Teachers' opinions of assessment results - Teachers' purposes for using assessment results
Teachers' views regarding the influence of institutions on assessment	<ul style="list-style-type: none"> - Teachers' opinions about institutional expectations regarding assessment - Institutional expectations of assessment - Influence of institution on teachers assessment perception and practices
Teachers' thoughts on the influence of the educational policy of the country on assessment	<ul style="list-style-type: none"> - Expectations of assessment policy from teachers - Influence of educational policy on teachers' assessment perceptions and practices

In the table above, themes that emerged from the analysis of data collected from various sources and the sub-themes belonging to each theme are presented. In the following parts, the emerging themes such as teachers' assessment perception, teachers' views about assessment technique, teachers' views concerning assessment results, teachers' views regarding the influence of institutions on assessment, and teachers' thoughts on the influence of the educational policy of the country on assessment are discussed in detail along with categories successively. While reporting the results, each sub-theme was presented with the help of a table for better understanding. Then, codes referring to striking results, viewed as crucial by the researcher, were explained under the tables with extracts taken from the obtained data.



*Figure 3.*Emerging Themes Regarding Teachers' Nature of Cognitions About Language Assessment

When the nature of cognitions of the teachers are examined before the RRD sessions, the area with the highest frequency of citation is their general perceptions about assessment. This theme was followed by themes that referred to teachers' views on assessment techniques, their ideas about the results of their assessment, and their views on the impact of the institution's and country's assessment policy on their assessment perceptions and practices.

4.2.1. Teachers' Assessment Perception

This emerging theme presents how participants perceived the assessment before the RRD sessions. Teachers' assessment perception theme appeared with six broad sub-themes such as teachers' definition of assessment, teachers' views on classroom assessment, teachers' purposes of assessment, teachers' views of ideal time for assessment, teachers' views on

effective assessment, teachers' initial perception of assessment and change in it over time. Excerpts illustrated how the teachers generally perceived assessment at the beginning of the research.

4.2.1.1. Teachers' Definition of Assessment

At the beginning of the study, teachers were asked questions about how they defined assessment to explore their assessment-related cognitions. On the basis of answers, it was found that the teachers described assessment in different ways. To put it roughly, all teachers defined assessment as administering exams especially written exams to students. This emerging sub-theme consists of findings referring to how teachers viewed and described the assessment before RRD sessions. Table 5 informs how teachers personally describe assessment.

Table 5.

Teachers' Definition of Assessment

Code	f	Code Description
Assessment as testing students' knowledge level	9	Participants explain assessment as a means of identifying students' learning levels.
Assessment as administering written exams	8	Participants imply that assessment is applying exams, especially written ones, to students.
Assessment as a way of evaluating teaching efficacy	4	Participants' definition of assessment is found to be about practices done to evaluate teaching by the teachers.
Assessment as a process	2	Participants explain how they see the assessment. Assessment as process-oriented.

When foreign language instructors were asked to define assessment in their terms, all of them described assessment as a practice through which teachers can determine whether students have learnt what was taught to them or not. To illustrate, when we examined how the teachers defined assessment in their own words, we found a similar pattern in all definitions. The excerpts below demonstrate these definitions, and the overlapping words are presented in bold.

*I can say that assessment is a **phase** at which the teacher can determine the **outcome**. In other words, assessment can be a **process** in which students' **learning levels** are **identified** through **exams** (pre-semi-structured interview, P2)*

*When I was asked the definition of assessment, a practice done in order to **discover** whether **students have learnt** what **we thought** to them or not comes to my mind. (Pre-semi-structured interview, P3)*

*Assessment is a technique of **specifying** what students have learnt and what they haven't learnt during a course. That is, assessment determines students' knowledge and specifies whether they **have learnt** what **I gave** them. I can define assessment like this. (Pre-semi-structured interview, P4)*

*When the assessment is asked of me, it also comes to my mind whether **I was able to teach** the subject effectively **to students**. That is, **determining** whether a teacher can influentially teach the issue comes to my mind. (Pre-semi-structured interview, P3)*

From the excerpts taken from participants' answers, it can be said that the teachers' word choice demonstrates that assessment is a means to diagnose whether students have understood the subject, attained the goals and the weaknesses that teachers have concerning teaching. It can also be stated that teachers perceive assessment as a way to evaluate their teaching effectiveness. Therefore, on the basis of the results, it can be indicated that participant instructors perceived assessment as a way not only to determine students' learning outcomes but also to explore their success in teaching.

4.2.1.2. Teachers' Views on Classroom Assessment

Participants were asked about what they knew regarding classroom assessment. As a result of the analysis of their answers in the interviews, it was found that there was a visual consistency between how the participant teachers perceived and defined the term assessment and how they perceived classroom assessment. We discovered that participant instructors detailed their perception of classroom assessment just as they previously did while describing the assessment. To go into detail, it was figured out that the teachers who did not refer to the concept of any improvement while defining the classroom assessment similarly could not stay away from the term, assessment, and emphasised assigning grades. That is, most teachers were found to associate classroom assessment with identifying whether students understood what was taught to them and giving grades accordingly. The excerpts illustrate how teachers perceived classroom assessment.

*I think of classroom assessment as a **practice** through which teachers can **map out** whether students have grabbed what was given to them and **identify students' degree of learning**. Classroom assessment is the **teacher's determination** of the **student's learning level** in the classroom. For example, I can say that the teacher **administers the students a quiz and grades** the performances of the students. (Pre-semi-structured interview, P1)*

Classroom assessment is a way in which a teacher can determine whether students have attained the objectives or not. (Pre-semi-structured interviews, P4)

The highlighted statements in the excerpts taken from the interviews exemplify how participant teachers perceived the classroom assessment. While the participant teachers were talking about classroom assessment, they could not go beyond revealing students' learning levels and associated classroom assessment with grading and administering exams as they did when defining the term assessment. The overlapping statements in the quotations show that they similarly perceived classroom assessment and assessment.

However, P3's answer differed slightly from the rest of the participants. When s/he was asked to mention how s/he perceived and defined the classroom assessment, s/he stated that classroom assessment was anything done to facilitate students' learning. S/he did not refer to any exam or a procedure which helps teachers identify students' competency levels. Instead, s/he focused on the function of classroom assessment by highlighting improving students' learning. The following excerpt illustrates his/her opinion:

Classroom assessment is an act or behaviour whose aim is developing students and facilitating their learning. Classroom assessment supports students' learning and increases teachers' teaching quality in assessment and evaluation practices. Assessment for learning. (Pre-semi-structured interview, P3)

As it can be understood from the highlighted parts in the excerpt above, one of the participants referred to development, supporting learning and teaching while defining classroom assessment. Instead of exploring students' current learning levels, we found that s/he perceived the classroom assessment as something aiming at uplifting the current status of students' learning and instructors' teaching to a better position.

4.2.1.3. Teachers' Purposes of Assessment

This sub-theme provides findings that refer to the participants' perceptions about assessment purposes. In other words, participants expressed their opinions about for what purposes assessment should be conducted before RRD sessions. Foreign language instructors also talked about their assessment purposes, which helped reveal their perceptions of assessment purpose(s). Table 6 provides teachers' overall perception of assessment purposes:

Table 6.

Teachers' Purposes of Assessment

Code	f	Code Description
Revealing students' learning levels	12	Teachers' purpose of assessment, checking student understanding.
Making decisions about students' success	11	Teachers' assessment purpose focusing on deciding on students' success based on obtained results
Identifying teaching efficiency	10	Teachers' assessment perception about identifying the effectiveness of their teaching.
Determining students' prior knowledge	5	Teachers' assessment purpose in terms of identifying students' current knowledge before the instruction, diagnostic purpose

Based on the analysis, we can say that teachers' perceptions about the aims of assessment aligned with their perceived definition of assessment. To illustrate, most participants were found to associate assessment with practices done by teachers to reveal students' learning outcomes. Also, teachers defined assessment as a means through which they could explore if the students achieved their goals or not. In light of the analysis results aiming to reveal their perceptions about the assessment purpose and their explored assessment perceptions, it can be said that the teachers' perceptions of assessment purposes are shaped by how they perceived the assessment. The following excerpts exemplify the participants' perceptions of the assessment purposes. The highlighted parts in the quotations show the parallelism between the participants' perceived definition of assessment and their perception of the general assessment purpose.

*In my opinion, the purpose of assessment is to **determine whether students have learned a topic** after teaching it (Pre-semi-structured interview, P2)*

*Assessment is **a way of determining to what extent students have learned the subject**. In other words, assessment is a practice applied **by teachers to check and determine students' understanding levels**. (Pre-semi-structured interview, p4)*

As it can be understood from the participants' explanations, teachers' most frequently stated aim of the assessment was to discover the learning levels of the students, which were also found to align with their assessment perceptions. As a consequence of the analysis, another result describing the participants' perceptions about the purpose of the assessment was also found to be related to determining the student's prior knowledge level. To be more precise, the participants, who claimed that one of the purposes of the assessment was to unveil

the learning levels of the students after the instruction, stated that another purpose of the assessment was to determine students' prior knowledge to reveal their existing knowledge before the instruction and plan the lesson accordingly. Based on this result, it can be deduced that although the assessment time is different, the assessment purpose perceived by the teachers is still linked to determining the knowledge level of the students. The following excerpts show teachers' views on this issue:

*I want to see **what students already know**. I must determine students' readiness to learn **to plan my lesson**. For the reason I mentioned, in my opinion, one of the primary purposes of assessment and evaluation is **to reveal students' existing knowledge**. (Pre-semi-structured interview, P1).*

*First, I should clearly **determine** how I should **plan the lesson**, what I will teach, and how I will teach in my mind. I should **identify my learners' prior knowledge** to learn to plan my lessons beforehand. **What do they know about the topic before they start the lesson** so I can help them construct the knowledge? I **need to determine this**. For **this reason, I assess and evaluate**. That is one of the purposes. (Pre-semi-structured interview, P3).*

As it can be understood from the highlighted parts in the excerpts above, the assessment purposes of the participants revolved around determining the level of knowledge that the students have. In addition, it informed us that instructors also assess their students to make decisions about their success. This result is related to the assessment perceptions of the teachers. In the perceived assessment definitions of the teachers, ideas such as determining the students' learning status, administering exams and making decisions about student success came to the fore. This result also revealed that teachers' perceptions of assessment affected their assessment practices. In other words, teachers' perceptions of assessment and thoughts about assessment purposes were in the same direction. This result can be said to be aligned with teachers' self-perceived assessment definitions. The following quotes exemplify the situation:

*We use the data obtained through exams we apply in the preparatory program to decide **whether students should pass or fail the preparatory class**. We also administer exams and get and use exam results to **decide on this issue**. (Pre-semi-structured interview, P1)*

*I also teach English in other undergraduate departments. Sometimes, I assess students in addition to mid-term and final exams. Most of the time, I **decide whether the student will be successful in the course** with those exams, but to be sure of myself, I assess my students again, except for exam times, **to make the right decision about their success**. For example, I observe them; I ask questions during the lesson. (Pre-structured interview, P4)*

When the excerpts are examined, we can say that teachers believed that assessment should be conducted to decide on students' success. To put it another way, it was found that teachers' assessment purposes were more student-oriented. In their statements related to the assessment purposes, no findings have emerged referring to a purpose such as assessing for improvement of the institution. On the contrary, more student-based goals were frequently mentioned.

4.2.1.4 Teachers' Views of Ideal Time for Assessment

Participant EFL instructors were asked to state their views on the ideal assessment time during the data collection. As a result of the analysis of the obtained data, different opinions emerged in terms of the appropriate time for assessment. This sub-theme presents the findings obtained from the analyses of teachers' views on the proper time for assessment. While some participants thought that assessment should be carried out simultaneously with the teaching and spread to the process, some participants stated that students should be assessed after teaching certain target subjects, just like in summative assessment. Table 7 details the sub-theme:

Table 7.

Teachers' Views of Ideal Time for Assessment

Code	f	Code Description
Assessing at the end of a teaching process	11	Teachers' perception considering the ideal time for assessment. Teachers thought that assessment should be done at the end of a certain period.
Assessing students before teaching.	3	Teachers' beliefs regarding the ideal time for assessment. Assessment should focus on identifying students' prior knowledge before teaching.
Spreading assessment throughout the process	2	Assessment should be continued together with teaching. It should be implemented throughout the process.

To go into detail, participants believed that assessing students must take place after spending a certain amount of time. To illustrate, teachers stated that students should be

assessed after teaching English to students for a certain period. Again, these findings show that participants' ideas about the ideal time for assessment align with their perceptions about the purpose of assessment. In other words, teachers, who stated that the purpose of the assessment was both to decide on the success of students and to determine the learning levels of students, said that the appropriate time for assessment should be, for example, at the end of a few units or the end of a semester, after a certain amount of time spent with teaching. The excerpts below, especially the highlighted parts, exemplify teachers' perceptions of the subject.

*I can only assess my students if I want. To assess my students, **I must teach something to my students**. As I move on to teaching, **as time possess**, I should assess my students **at the end**. (Pre-semi-structured interview, P2).*

*The most **appropriate time** for assessment is the **end of the unit or month**. To illustrate, **I assess my student after I teach and finish the lesson**. (Post-semi-structured interview, P4).*

*Assessment should be done in the **middle of the semester or at the end**. That is, an assessment should be done **after teaching some units**. (Pre-semi-structured interview, P3).*

As seen from overlapping words included in participants' explanations, teachers believe that students should be assessed only after they are taught some subjects. In conclusion, it could be claimed that participants' ideal time for the assessment is at the end of a teaching process.

Secondly, another ideal time, which the participants clearly emphasized, was found as the beginning of the semester. As can be understood from Table 7, teachers stated that they aimed to determine students' current level of present knowledge to determine their level of readiness to learn and to plan their lessons and curriculum effectively at the beginning of the year. All these emerging findings show that the teachers' assessment perceptions, perceptions of assessment purposes, and ideas about the ideal time for assessment are related to each other. To explain more, the teachers who stated that one of the purposes of assessment that they considered important was to identify students' prior knowledge naturally presented the beginning of the semester as an appropriate time for assessment. The quotations exemplify this finding:

*The **appropriate time** to assess is also the **beginning of the year**. Because at the beginning of the year, I would like to assess students and see **what they already have in their minds**. I may change my curriculum according to the data I have obtained. For example, I may need to make it easier, or I generally know how to conduct my lessons. (Pre-semi-structured interview, P2).*

*One of the appropriate times for assessment is **the beginning of the semester** or the **beginning of a unit** because we can shape our teaching according to the assessment results. Then, we won't teach a difficult topic well above the students' level or waste time explaining things that students already know. (Pre-semi-structured interview, P2)*

As a result, the suitable time for assessment was found as the end of a certain period of time spent on teachings, such as the end of the unit or the end of the semester, to identify the learning outcomes of the students and the beginning of a semester or a unit to plan the instruction effectively.

4.2.1.5. Teachers' Views on Effective Assessment

EFL instructors shared their opinions about the features of effective assessment. Based on the answers given by the teachers, findings referring to their views on how an effective assessment was gathered under the title of this sub-theme. That is, this sub-theme included in Table 8 provides findings concerning teachers' views on specific characteristics of an effective assessment.

Table 8.

Teachers' Views on Effective Assessment

Code	f	Code Description
Practical	11	Assessment should be easy to apply, score and grade. Administering it should be easy for teachers and students.
Objectivity	10	According to the teachers, assessment should be fair, equal and objective. Subjectivity and bias should not be included.
Assessment focusing on the process	5	Assessment should be process-oriented rather than product oriented. Teachers need to conduct an assessment and teach simultaneously.
Economic	5	Teachers' perception of how an ideal assessment should be. Assessment should be economical in terms of time and effort.

As can be seen from Table 8, the first feature frequently uttered by participants emerged as objectivity. The common result of the analysis of teachers' statements is that assessment and evaluation should be equal and objective for all students, independent of teachers' prejudices and subjective opinions. Participants advocated grading in the light of criteria valid for all students rather than teachers' personal decisions. Looking at the excerpts below and the highlighted parts accompanying them, teachers often mentioned grading while discussing effective assessment. Based on this, it can be said that grading is dominant in determining teachers' perception of assessment and its quality. The excerpts explicate participants' views on effective assessment.

Effective assessment should be objective. Students should be assessed and evaluated based on their grades from the exams, not their thoughts. Teachers should give students the grades they deserve and be fair in grading. Thus, an effective assessment can be performed. (Pre-semi-structured interview, P3).

Above all, an effective assessment must be objective. Students should be assessed, and their performance should be graded in the light of predetermined criteria independent of personal considerations and applied to all students. (Pre-semi-structured interview, P4).

Secondly, teachers stated that an effective assessment is a practice through which students are assessed not only by considering their final grades at the end of a certain period, such as a semester, but also taking their development and performance throughout this period into account. The excerpt exemplifies the finding below.

*Assessment should be seen as a **process**. Assessment practices should be **distributed throughout the process**. Students shouldn't be assessed **considering a result**. The focus should be **on the student's development throughout the process**, not on the grade they receive at the end. (Pre-semi-structured interview, P1).*

To sum up, when we examined how the teachers perceived an effective assessment in their words, two typical characteristics helped us understand what they thought about it. As indicated previously, teachers believed that students should be assessed in a process-oriented and objective way.

4.2.1.6. Teachers' Initial Perception of Assessment and Change in it over Time

As a result of the analysis, this study has identified a new sub-theme called teachers' initial perception of assessment and change in it. This sub-theme gives information about teachers' perceptions when they first started the profession and how these perceptions developed over time and with experience.

Table 9.

Teachers' Initial Perception of Assessment and Change in it over Time

Code	f	Code Description
Assessment Technique	9	This code gives information about change in teachers' cognitions and practices of assessment techniques which shifts from traditional to alternative techniques.
Assessment Type	8	This code gives information about change in teachers' cognitions and practices of assessment type in time which shifts from diagnostic and formative to summative assessment.
Assessment Frequency	6	This code gives information about change in teachers' both cognitions and practices of assessment frequency in time which is from more to less frequent.
Assessment Time	4	This code gives information about the change in teachers' cognitions and practices of assessment time in time and experience.

When the participants were asked about their initial perceptions of assessment and whether there was a change in it over time, the teachers expressed a variety of perspectives. As can be understood from Table 9 above, these views are related to the assessment technique, assessment time, assessment frequency and assessment type. The following section explains the teachers' initial perceptions of these identified issues and their change over time.

Teachers first mentioned their initial ideal assessment technique and how it changed over time. They indicated that at the beginning, they favored sit-down written paper and

pencil tests administered at the end of a certain period. Later on, as they gained teaching experience, they realized that only written exams were insufficient. Therefore they started to use contemporary and alternative assessment techniques. While talking about this issue, interviewees said:

At the beginning of my career, I used to believe that students must show their knowledge on a written exam paper. For this reason, I used to adopt only written paper and pencil exams. As I continued in my profession, there has been a change considering the preferred assessment technique. Namely, I started to use contemporary ones. I mean alternative techniques such as the portfolio or peer assessment (Pre-semi-structure- interview, P2).

My previous perception of assessment was that teachers could effectively assess and evaluate their students at the end of term through sit-down written exams called mid-term and final exams. Over time, I realized that only written exams administered 3-4 times a year were insufficient to decide students' English proficiency. For example, I am trying to apply the performance assessment as far as I know. (Pre-semi-structured interview, P3).

This result indicates that as teachers experienced teaching English, they realized the missing parts of traditional assessment techniques and turned their attention to alternative assessment to make up for the shortcomings of traditional assessment techniques.

The following result related to this sub-theme is concerned with assessment time. Some of the teachers interviewed remarked that although they were in the beginning in favor of assessing students at the end of a certain time, they realized the effectiveness of assessing students during their teaching process. To illustrate, a shift was figured out in participants' perceptions from product-oriented assessment to process-oriented one. This finding is exemplified by the excerpts below.

At the beginning of my teaching career, I supported just mid-terms administered at the end of the semester. At the beginning of my teaching career, assessing students at the end of the semester was enough. So by assessing students at the end of a few units or the end of the semester, I felt like I was missing something. That's why I decided to assess students more frequently throughout the semester rather than only at the end of the year. (Pre-semi-structured interview, P2).

It was also determined based on the results that there was a change over time related to the assessment frequency of the participants. Teachers initially assessing their students less frequently at the end of a certain time explained that they started to assess them more frequently over time. This result is in line with teachers' changing perceptions considering the appropriate time for assessment. Teachers who assessed their students less frequently at the end of a period of time increased their frequency of assessment and started to assess their students by focusing on the process throughout the term. This result is apparent in the excerpt below taken from the interview.

*When I first graduated from the university, I used to think that assessing **students depending on exam time** determined by the institution would be enough. According to the exam timetable of the school, we used to administer three written exams at the beginning, in the middle and at the end of the semester successively. As I **became more experienced**, I realized that having an exam mood is better for students also, I can be more effectively informed about students' levels. Therefore, I **increased the frequency of exams**; I am **assessing my students more frequently** now. (Pre-semi-structured interview, P4).*

In addition to the first three findings, it was also found that based on the analysis, teachers were found to adopt summative assessment in their assessment practices at the beginning of their careers. Then, they reported a change from summative to formative assessment. In other words, throughout time and experience, teachers became interested in monitoring student learning to provide students with continuing feedback for enhancing their learning rather than only showing students their final understanding of a particular topic. The excerpt below illustrates how this change in assessment type was realised.

*I was more focused on **summative assessment** when I **first started teaching**. I was only interested in the students' grades on the exams, and according to them, I would have an idea about the student's success. I used to give grades then I used to continue the lesson. In the following exams, I could not encounter any improvement. Then, I became **more inclined to prefer formative assessment in my practices** believing that I could **improve the quality of students' learning**. (Pre-semi-structured interview, P2).*

These results suggest a hidden association among teachers' perceptions of assessment time, assessment technique, frequency and type. Both instructors' views of these concerns in their early years of teaching and their perceptions of the topics above, which were claimed to change in time and with experience, exhibit a particular pattern and influence one another's

nature and appearance. For instance, the analysis revealed that teachers, over time, started to believe that teachers should adopt formative assessment and increase the number of assessments. As a result of these, it also resulted in the perception that teachers should assess their students more frequently by focusing on processes and using alternative assessment techniques.

4.2.2. Teachers' Views about Assessment Technique

Throughout the interviews, teachers were provided with questions about their knowledge, preference for assessment techniques, and their purposes for implementing them in their assessment practices. Also, they were asked questions to find out the reason behind their preference of the assessment technique(s) they claim to use, their opinions about the assessment technique and their alternative assessment knowledge.

Data elicited from the transcripts of semi-structured interviews regarding the theme 'teachers' views about assessment technique,' five sub-themes such as teachers' assessment technique knowledge and preference, teachers' purpose of using assessment technique, reason(s) behind teachers' preference for assessment technique, teachers' perception of effective assessment technique and its characteristics and teachers' perception of alternative assessment. These sub-themes are explained in the following sections with the help of tables and excerpts taken from the original data.

4.2.2.1 Teachers' Assessment Technique Knowledge and Preference

During the interviews, teachers were asked some questions to explore their knowledge and specific assessment technique preferences. As a result of the analysis, this emerging sub-theme consists of three codes: knowing and practicing written exams, limited knowledge of assessment techniques, incomplete knowledge and practice regarding alternative techniques. Table 10 summarises this emerging sub-theme:

Table 10.

Teachers' Assessment Technique Knowledge and Preference

Code	f	Code Description
Knowing and practicing sit-down written exams	8	Teachers' current knowledge and practice considering assessment technique: written sit-down exams.
Incomplete knowledge and practice regarding alternative techniques	6	Regarding assessment technique knowledge and practice, teachers' opinions about their competency considering this issue, having incomplete knowledge of assessment techniques.
Limited knowledge of assessment techniques	5	Perception of having limited knowledge of kinds of assessment techniques.

From Table 10, one can see that a common point in the answers concerning assessment techniques known by the participants is that they show written and oral exams as the assessment techniques they learn. Also, it was revealed that there was a visible consistency between participants' assessment knowledge and the assessment techniques they preferred in their actual classroom practices. To illustrate, most participants also used written exams while assessing their students in their classrooms. Additionally, teachers were observed in their classes, and it was seen that written exams called quizzes are applied while assessing their students. Therefore, we can say that teachers' knowledge matches their actual classroom practices. The excerpts below provide evidence about participants' knowledge and preference for assessment techniques. Also, words presented in bold transmit the teachers' answers related to the finding.

Regarding assessment techniques, I know only written paper and pencil exams and oral exams. We administer these exams periodically, and I know them (Pre-semi-structured interview, P4).

We administer placement tests at the beginning of the semester to put students into groups appropriate for their levels; I know oral exams. I know paper and pencil exams such as midterm and final exams. I know them, and I use them. They serve effectively in achieving my goals. (Pre-semi-structured interview, P3).

I always try to assess my students. I ask verbal questions; I administer quizzes. I use written exams and oral exams more frequently in my class practice. (Pre-semi-structured interview, P1)

The next code emerged as a result of the analysis concerned with participants' weakness in knowledge of different assessment techniques. From the answers given by them, it was seen that they either directly reported not having much knowledge about assessment techniques or they just showed written and oral exams as techniques that they knew. When considering the kinds of assessment techniques, we can say that teachers' knowledge concerning assessment techniques is limited. Instructor P2's explanations are given to exemplify the result:

Summative assessment is a technique, formative assessment is a technique. Although I try to use formative assessment techniques, I use summative techniques to meet the expectations of my institution. (Pre-semi-structured interview, P2).

*Frankly, I only know a little about assessment techniques **other than written and oral exams**, and I **don't apply various techniques** because I **do not know them**. I've heard of **alternative assessment techniques** in my undergraduate education, but I **need to learn them better**. (Pre-semi-structured interview, P2).*

From the explanations above, it is clear that participants confused assessment techniques with types of assessment which might be interpreted as a weakness in participants' assessment technique knowledge.

In addition to these, one of the participants reported knowing and preferring alternative assessment techniques in their practices. P2 explained: "I mostly focus on using portfolio and classroom observation in my classroom assessment practices because the portfolio is flexible." S/he explained the reason why s/he used portfolios in assessing their students. Also, participant P1 reported using classroom observation as an assessment technique in their practices by saying, "I also assess my students based on classroom assessment as well." After the interviews, participants were observed in their actual classroom. However, they were not observed while applying the alternative techniques they claimed to use. Then teachers were asked some questions in the post-semi-structured interview session held after observations. Based on the data from such procedures, it was found that although some participants claim to know and use alternative techniques, their knowledge and applications were incomplete.

These results indicate that instructors' assessment technique knowledge and usage were limited to traditional assessment.

4.2.2.2 Teachers' Purpose of Using Assessment Technique

During the interview, participants mentioned why and for what purposes they implemented assessment techniques. Their answers identified various purposes for using assessment techniques, such as checking students' understanding, making decisions about student success, and tracing students' development in time. Table 11 below presents codes constituting this emerging sub-theme:

Table 11.

Teachers' Purpose of Using Assessment Technique

Code	f	Code Description
Checking students' understanding	14	Teachers' purposes in using assessment techniques which is finding out students' understanding levels.
Making decisions about student success	11	Teachers' purposes of using assessment techniques are making some decisions on students' success, such as passed/failed decisions.
To find out teacher's teaching efficiency	8	Teachers' self-perceived purpose of using assessment techniques – evaluating the effectiveness of teaching.
Tracing students' development in time	2	Teachers' purpose of using assessment techniques to trace students' development throughout the teaching process.

As stated before, this emerging sub-theme informs about the purposes for which teachers use assessment tools or techniques in their assessment and evaluation practices. Firstly, participant teachers mostly use assessment techniques to check students' understanding. The excerpts exemplify the situation:

I use assessment techniques to determine whether my students have learned the subject. I use assessment techniques to see whether students have attained the goals at the end of a unit. (Pre-semi-structured interview, P1).

I assess my students at the end of units or a month or a semester to see whether students can learn or not or whether students can grab what I intended to give them. I use assessment techniques to determine whether students learned target information. (Pre-semi-structured interview, P4).

I use assessment techniques to check out students' understanding. (Pre-semi-structured interview, P3).

We encountered almost similar expressions when we examined how teachers expressed their purpose for using assessment techniques. Based on the overlapping phrases presented in the excerpts above, we can say that teachers' first and foremost aim was to shed light on whether students could attain the pre-determined goals or not. Furthermore, it can be said that teachers focus on students' learning gains, namely products, at the end of a certain period spent teaching. Moreover, this result aligns with how teachers perceived and defined the assessment. To be more precise, when the teachers' responses were considered, their definition of assessment and self-perceived purpose of using assessment technique, the phrase "finding out students' understanding level" was found to be common in answers to both topics. This result suggests that teachers' perception of assessment determines their purpose in implementing assessment techniques.

What is more, participants reported that they used assessment techniques to make decisions about students' success. This finding is also found to be in line with results regarding teachers' perceived definition of assessment. Considering the results that emerged concerning both topics, teachers mentioned making decisions on student success. To explain further, while the teachers defined assessment as a practice conducted to determine student achievement, they also emphasized a similar point while explaining their purpose for implementing the assessment technique. Based on these interconnected results, we can say that teachers' perceptions of assessment still determine their goal of using assessment techniques. The excerpt illustrates how teachers aim to make decisions on students' success by using assessment techniques:

*I use assessment techniques to say whether students are successful or unsuccessful at the end of the process. That is, in preparatory classes, we subject students to quizzes and mid-terms many times. By looking at the average grades the students received from them, we **decide whether they passed/failed the preparatory classes**. (Pre-semi-structured interview, P3).*

4.2.2.3. Reasons Behind Teachers' Preference for Assessment Technique

During the interviews, participants were asked to explain the reason(s) for adopting assessment technique(s) that they implemented in their practices. They proposed many reasons, such as practicality, the influence of technique on students, institutional requirements and the benefit of the technique. In Table 12, an overview of participant teachers' perceived reasons behind their assessment technique selection is given:

Table 12.

Reasons Behind Teachers' Preference for Assessment Technique

Code	f	Code Description
Institutional requirements	9	Selecting assessment techniques based on institutional needs and requirements.
Practicality	8	Choosing any assessment technique since it is easy to administer, score and grade.
Benefits of the technique	4	Choosing an assessment technique or tool for the benefits that it provides both to students and teachers.
Influence of technique on students	3	Choosing an assessment technique for its positive influence on students.

We can see that participants mostly referred to practicality while explaining the reasons for selecting assessment techniques for their practices. They stated that they preferred written exam since it is practical for both teachers and students. Students claim that students are familiar with this type of assessment technique, and as a result, they do not have difficulty. As for the teachers, they claim that written exams are more advantageous since they are easy to administer, score and grade. Considering the teachers' own words, we can say that teachers first consider practicality when selecting the assessment technique. Moreover, based on the results, it was revealed that teachers mostly chose written exams for the classroom practices. This finding is in line with the previously found result showing written exams as the most frequent assessment technique preferred by the teachers. The excerpts below show teachers' views on practicality as a reason behind their assessment technique selection:

*Our classes are crowded. We have many classes to teach, and written **exams** can be **administered easily**. They are also **economical** in terms of **time, effort and expense**. (Pre-semi-structured interview, P3).*

***Written exams** don't require too much preparation; they **are practical and usable** (Pre-semi-structured interview, P1).*

The result concerning the fact that teachers act in parallel with institutional requirements when deciding on the type of assessment technique can be reached. In other words, the teachers emphasized that the institution they work for requires them to implement written exams in their assessment practices. Therefore, they frequently choose written exams. This result suggests that the expectation of the institution from teachers is one of the determinant factors in teachers' selection of assessment technique(s). The sentences quoted from the statements of the teachers in the interviews exemplify this result:

Circumstances, opportunities given by the institution, the limited time that I have, and legislation provided by our SFL led me to adopt written exams as assessment techniques. (Pre-semi-structured interview, P1).

Since my institution asked me to administer written exams called mid-terms and quizzes, I adopted sit-down written paper and pencil tests. The school asks us for a written document as proof of the grade we give to the student, so it wants us to do our midterms and finals as written exams. (Pre-semi-structured interview, P3).

Moreover, the influence of the assessment techniques on students' learning effort and performance was shown as a factor determining their assessment technique preference. From the answers given, we understood that teachers want to see a positive influence of the assessment technique on their students' learning. Accordingly, teachers said they mostly prefer written exams since they were thought to motivate the students. According to teachers, the students also select written exams as an ideal assessment technique. Then, based on the explanations, we can suggest that teachers consider students' reactions while making decisions on assessment technique preference. The excerpts show how teachers consider the effect of the technique on the student when choosing an assessment technique.

I usually use sit-down written exams as an assessment technique because it not only motivates the students but it also conveys the fear of grades to students (Pre-semi structured interview, P2)

I choose written and oral exams as assessment techniques in my practice because my students value them. They think that they can show their knowledge practically and effectively through written exams. (Pre-semi structured interview, P4).

In summary, teachers almost exclusively referred to written exams while answering all questions that aimed to reveal their reasons for choosing their assessment technique. The fact that teachers' answers show a visible pattern around the written exam technique indicates that written exams are dominant in their assessment technique knowledge and practices. The next part provides emerging results concerning teachers' ideas on effective assessment techniques and its characteristics.

4.2.2.4. Teachers' Perception of Effective Assessment Technique and Its Characteristics

This sub-theme provides results that refer to teachers' opinions about the best assessment technique and its general characteristics. During the interviews, participants were asked to identify the best assessment technique and express how it should be. From the

explanations given by the participants, teachers showed written exams as the best assessment methods and practicality, validity and reliability were presented as general characteristics of the written exams, which turn them into the best assessment techniques. Table 13 summarizes the sub-theme:

Table 13.

Teachers' Perception of Effective Assessment Technique and Its Characteristics

Code	f	Code Description
Written exams for its validity	9	Teachers perceive the written exams as best method because of its content validity
Written exams for its practicality	7	Teachers perceive written exams as the best method for the practicality provided by it
Written exams for its reliability	4	Teachers perceive the written exams as best method because of its content validity, not including measurement error.

On the basis of analysis, we can say that teachers regard the assessment technique that they mostly use in their practices namely written exams as best assessment technique. This result suggests that there is an association between instructors' opinions of best assessment technique and their proposed assessment technique usage.

Furthermore, most of the participants provided with some specific characteristics while they were explaining their opinions about best method. As a result of the analysis, instructors' explanations concerning general characteristics of the method showed consistency with their statements about reasons behind their assessment technique preference. For instance, P1 stated that s/he mostly used written exam in his/her assessment practices because of its practicality. Also, s/he reported that the best assessment method was written exam. Additionally, s/he provided with reasons why s/he perceived written exam as the best assessment technique. Following excerpt will exemplifies the situation:

According to me, sit-down written exam is the best assessment technique because it is practical for me to apply and grade it. Also students are familiar with this technique and they are get used to be assessed through it. Additionally, students take written exams serious and they are motivated to study. That's why, written exam is best technique for me (Pre semi-structured interview, P1).

Additionally, teachers presented objectivity, content validity, reliability and practicability as factors making written exams best techniques. While the teachers were talking about the features of the best assessment technique, they also referred to the written

exams and stated that thanks to the written exams, they were able to ask questions assessing all objectives at once, both reliably and practically. The following excerpt together with some words in bold illustrate the situation:

*Written exams such as quizzes are **the best** assessment techniques. Because they are **practical** and most of the time they are **objective**. Through written exams, we can **ask many questions** at once about **various objectives** that we set. I mean, their **content validity** is high and they are also **reliable**, I think. (Pre semi-structured interview, P3).*

4.2.2.5. Teachers' Perception of Alternative Assessment

During the interviews, participants also were provided with some questions about alternative assessment techniques. This current sub-theme, which emerged as a result of the analysis of teachers' statements regarding alternative assessment techniques, first presents teachers' views on alternative assessment techniques. This is followed by results referring to teachers' knowledge of alternative assessment techniques and teachers' implementation of those techniques in their practices successively. Table 14 details the codes of this emerging sub-theme:

Table 14.

Teachers' Perception of Alternative Assessment

Code	f	Code Description
Teachers' knowledge of specific alternative assessment techniques	8	This code refers to findings showing teachers' knowledge of kinds of alternative assessment techniques. Alternative assessment techniques known by the teachers.
Teachers' implementation of alternative assessment techniques	8	This code shows to what extent teachers implement alternative assessment techniques in their actual practices.
Teachers' opinions of alternative assessment techniques	4	This code describes what teachers perceive namely think about alternatives assessment techniques- their point of views regarding alternative assessment techniques.

Based on the answers, we can indicate that participants have different ideas on alternative assessment techniques. It might be resulted from how they perceived the assessment and their assessment purposes. To go into detail, first, teachers reported seeing alternative assessment techniques as useful and a supporter of teachers in raising the quality of students' learning and teachers' teaching. P2 elaborates on the situation:

*I think **alternative assessment techniques**, such as the portfolio, **provide useful opportunities** for us. These assessment techniques **help us to follow our students***

*constantly and to **discover their progress and deficiencies**, if any, before it's too late. (Pre-semi-structured interview)*

On the basis of this emerging finding, we can say that the participant believes that assessment is a part of teaching and learning and it can support them.

Next, some negative feelings and thoughts regarding alternative assessment techniques were found out. To illustrate, some of the teachers reported that these techniques were too demanding for them. They required too much time and effort; that's why the teachers preferred traditional techniques like written and oral exams. This finding is exemplified the situation as follows:

*I think these alternative assessment techniques **complicate our work**. When I was a secondary school teacher, we had a very intense curriculum. Naturally, **we would devote our time and energy to teaching**. Instead of **dealing with these techniques**, we would immediately apply the written exams and continue teaching in the other course. (Pre-semi-structured interview, P4).*

Additionally, teachers reported that they do not exactly remember those techniques exactly. Actually, participants mentioned only portfolio when they were asked alternative techniques. Considering the various kinds of alternative assessment techniques, we can say that teachers' perceptions of alternative assessment techniques are limited to portfolio.

Moreover, teachers reported that they do not use alternative assessment techniques in their practices at all since they could not remember how to apply them. During the observation, their statements were confirmed. That is, they were not observed implementing any alternative assessment technique. This stated lack of practice also suggest that what teachers believe, know and do considering alternative assessment technique can be limited. The excerpts explain the result referring to teachers' lack of implementation of alternative assessment techniques:

*I remember that we learnt some alternative assessment techniques when we were at university but **I cannot remember these techniques exactly I forgot them so I can't use alternative assessment techniques** in my practices. (Pre-semi-structured interview, P1).*

*There are various kinds of alternative assessment techniques. **I heard** them while I was taking my bachelor degree. **May be**, I am using **unconsciously** and I am not aware of it.*

There are peer assessment and self-assessment. However, I don't know how to use them. (Pre-semi-structured interview, P4).

4.2.3. Teachers' Views Concerning Assessment Results

During the interviews, the participants were asked questions about assessment results. This theme, which emerged as a result of the analysis of the answers given, is concerned with information about teachers' perceptions about assessment results. Go even deeper, two sub-themes such as teachers' opinions of assessment results and teachers' purposes for using assessment results constitute the current theme. These sub-themes are explained in detail in the following sections.

4.2.3.1 Teachers' Opinions of Assessment Results

This emerging sub-theme reports results that refer to the opinions of the teachers regarding assessment results. On the basis of findings obtained from the statements of teachers, we can say that they generally have positive ideas towards assessment results. Firstly, teachers reported that assessment results were highly crucial for them because they perceive them as a useful means for achieving many goals such as evaluating their teaching efficacy, checking students' understanding and making decisions on students' success and planning instruction. Secondly, participant teacher reported that s/he viewed assessment results as helpful for themselves in their profession. Namely, assessment results help them adapt their lesson plans, activity and materials. Table 15 shows codes included in this emerging sub-theme:

Table 15.

Teachers' Opinions of Assessment Results

Code	f	Code Description
Helpful in achieving goals	8	Teachers perceive assessment results as a helpful means for attaining goals.
A complement of teaching	2	Teachers' thoughts and opinions regarding assessment results, perceiving assessment results as complement of teaching.

It is clear that teachers are trying to reach many of their teaching-related goals thanks to the assessment results. When the expressions of the teachers are examined, they perceive assessment results as an important source that sheds light on their way in terms of self-criticism of their teaching, determining the students' learning status, and making necessary adaptations in lesson plans and even in the curriculum for their further teaching practices. To

illustrate, we can say that since assessment results are useful to the teachers, therefore they are perceived as significant by them. The following excerpts especially highlighted parts explain teachers' views on assessment results and reasons why they perceive them as crucial.

Assessment results are important for me because thanks to assessment results, I can see my capabilities and strengths as teachers. (Pre-semi-structured interview, P4)

Assessment is important for me because I can check whether I am successful as a teacher or not by looking at the assessment results. (Pre-semi-structured interview, P1)

Assessment results are very important for me because by analysing them, I can check my students' understanding levels concerning the subjects I taught to them. (Pre-semi-structured interview, P3)

Assessment results are highly important for me because I plan my next instruction, my lesson by taking assessment results into consideration. (Pre-semi-structured interview, P2).

4.2.3.2 Teachers' Purposes for Using Assessment Results

Another sub-theme that makes up this emerging theme is "Teachers' Purposes for Using Assessment Results". It presents findings referring to teachers' purposes of using assessment results. On the basis of results, we can say that teachers utilise assessment results for various purposes such as using assessment results for checking effectiveness of teaching, exploring students' competency levels, deciding on students' success and planning further instruction. Table 16 illustrates the overall codes constituting this emerging sub-theme:

Table 16.

Teachers' Purposes for Using Assessment Results

Code	f	Code Description
Exploring students' competency level	17	Purpose of utilising assessment techniques for discovering students' understanding levels.
Deciding on students' success	15	Use of assessment techniques for making decisions about students' success / to pass or fail
Checking effectiveness of teaching	13	Purposes for implementing any assessment technique. Using assessment techniques for evaluating teachers' instruction to see it is effective or not.
Planning further instruction	4	Purpose to use some kinds of assessment techniques. Based on results obtained through them teachers can make some necessary adaptations on their lesson plans, activities and materials.

As can be seen from Table 16, it can be said that participants utilize assessment results for various purposes. According to the results, the most frequently expressed purpose of utilising assessment results by the participating teachers is to check students' learning outcomes. In light of the results, teachers give importance to whether students learn the subjects that are being taught to them. In order to determine this, teachers primarily use the assessment results for this purpose. This result is parallel with a previously revealed finding in this study. To elaborate, when the teachers were talking about the purposes of assessment, they first talked about exploring whether the students learned the target subjects or not. Likewise, this finding indicates that teachers pay attention to their students' understanding levels not only is specifying their assessment purposes but also determining their purpose for utilising assessment results. Regarding checking students' understanding levels, participants explained the situation:

I use assessment results in order to find out whether our students were able to learn mid-term subjects or not. (Pre-semi-structured interview, P3)

Later, it was found that one of the purposes of teachers to use assessment results was to determine how effective and beneficial they are to students as teachers and to make a self-evaluation in terms of teaching skills. In other words, teachers evaluate themselves by looking at the assessment results. Based on the statements of the participants, it can be said that, by analysing their assessment results, teachers explored the strengths and especially weaknesses related to their teaching and made the necessary modifications for their following teaching practices in order to improve their teaching. Concerning checking teaching efficiency, teachers illustrated the situation:

I use assessment results to check whether I was able to teach or not." When I realized that I have been able to teach the subject, I feel happy. (Pre-semi-structured interview, P3).

I use assessment results **to evaluate myself**. Based on assessment results, **I can define my weaknesses and strengths**. (Pre semi-structured interview, P2)

I use assessment results in order to evaluate myself. I use assessment results in order to answer these questions such as Are problems related to the students' inefficiency resulted from me? or Did I behave wrongly? (Pre-semi-structured interview, P1)

What is more, almost all participants stated that they made decisions about students' success by using assessment results. In other words, just as teachers get feedback about their

teaching effectiveness by looking at the assessment results, they also obtain information about the students and then use it, that is, the assessment results, to make decisions about the success of the students. The following excerpts explain the finding:

By looking assessment results, I make decisions about whether students can pass the prep class or not. I use assessment results while I am deciding on whether students should pass the prep class or fail. (Pre-semi-structured interview, P3)

I use assessment results in order to decide on whether students are successful or not from the preparation program. (Pre-semi-structured interview, P1)

Last but not the least, on the basis of obtained results, it was revealed that participants also attempted to plan or reshape their instruction by modifying their lesson plan and curriculum. Teachers also were found to either completely replace or adapt the activities and materials used in order to develop their further teaching practices. The excerpts clarify the situation:

I use assessment results in determining the content of my further lessons. What should be taught to the students? I try to find answer of this question. (Pre-semi-structured interview, P2)

I use assessment results in determining what should be changed and then I act accordingly. (Pre-semi-structured interview, P4)

Overall, these results indicate that teachers' purposes of using assessment results are in line with their assessment purposes in general.

4.2.4. Teachers' Views Regarding the Influence of Institutions on Assessment

This emerging theme conveys findings related to the impact of the institution in which teachers work on their assessment knowledge and practice. EFL instructors reported that institution had a significant role in their assessment knowledge and practices. The expressions provided by the teachers concerning institution and assessment were gathered under four sub-themes. The first one which is teachers' opinions about institutional expectations regarding assessment provides findings about what teachers' think about and evaluate what their institution expects from them to do regarding assessment. The next sub-theme includes assessment related institutional expectations of assessment. Finally, findings referring to influence of institution on teachers' assessment perception and practices are given. Each sub-theme is detailed in the following parts.

4.2.4.1 Teachers' Opinions about Institutional Expectations Regarding Assessment

This sub-theme presents results that refer to teachers' opinions about assessment-related expectations of their institutions from them. Based on the teachers' own words, it was revealed that while some of them were satisfied with the assessment policy of the institution they work for, some of the teachers did not like what was expected of them regarding the assessment. Teachers see the institutional assessment expectations as useful, in line with their assessment ideas, workload and useless respectively. Table 17 below provides an overall picture of this emerging sub-theme:

Table 17.

Teachers' Opinions about Institutional Expectations Regarding Assessment

Code	f	Code Description
Useful	6	How teachers see institutional expectations of assessment- their thoughts are parallel with what their institutions expect from them
Workload	4	The teachers are not satisfied with the institutional expectations regarding assessment. They do not appreciate it and perceive them as a workload
Useless	3	Teachers' do not appreciate institutional assessment-related expectations. According to them, they do not make any sense and are useless.

When it is examined, firstly, the teachers think that institutional expectations are beneficial. Previously, according to teachers, it was found that assessing students with written exams in the middle and at the end of the semester is sufficient for an effective assessment. For this reason, it was revealed that they appreciated the institutional assessment-related expectations. It is clearly seen that this result is in line with the findings showing the opinions of the participants about the assessment technique and the ideal time for assessment. Teachers, who assess students' learning outcomes with written exams at the end of a certain time spent with teaching, perceive these expectations as useful as the expectations of the institutions they work for do not contradict their own views on these issues. Using the excerpt below, one of the participants explains the aforementioned situation:

*In my opinion, **expectations** of our **administrators** from us are **beneficial** and **supportive** because they want us to **determine whether students have learned** or not. **Teach** the subject first, then **test** whether the students have learned that subject or not. It becomes like a mutual relationship between teaching and assessment. I think both of them support each other in this way. (Pre-semi-structured interview, P4).*

Nearly, half of the participants noted that institutional expectations of assessment were not enough to improve students' learning. Some of the participating teachers believed that the expectations of the institution were limited in terms of assessment frequency, therefore, the data obtained through these applications will not improve teaching. We can say that this idea of the participants is similar to the result referring to their perceptions of assessment that change over time. Some of the participants stated that they believed that assessment should be done at the end of the year in the beginning of their careers, but these thoughts developed over time and they started to think that assessment should be done frequently. This result indicates that teachers developing their instructions by utilising assessment results believe that only applying two exams throughout a semester in different time intervals are not sufficient for them to get feedback and make necessary adaptations in their instructional practices, therefore, the institution's assessment expectations are perceived by teachers as useless. The extracts explain the situation:

*Expectations of our institution considering assessment **limit the learning and don't help us improve our teaching**. They do not provide room for **development**. Because their number is only two one of which is a mid-term and the other one is a final exam. In order to take lessons from them, their number should be multiple. That's why, our institution's expectations **do not make sense to me** (Pre-semi-structured interview, P1).*

The last but not least, since the teachers do not see the assessment expectations of the institution as useful and do not believe that they will provide any benefit even if they fulfil them, they see these expectations as a workload. The following excerpt also exemplifies the situation:

*As I said before, **meeting such expectations is both a waste of time and a workload for me, since it does not provide any benefit**. Administering written exams to 100 students, then grading them one by one... I think that **it complicates our job**. (Pre-semi-structured interview, P1).*

As a result, these findings show that teachers have various ideas about the expectations of their institutions regarding assessment. These ideas were found to be connected with the findings obtained from teachers' own words concerning the ideal time for assessment, and assessment purposes.

4.2.4.2. Institutional Expectations of Assessment

While the previous sub-theme presented teachers' opinions concerning the institutional expectations related to assessment, the current sub-theme provides information about what these institutional expectations regarding assessment from the teachers are. To summarise roughly, these institutional expectations were found as expectations about the assessment time, expectations about the assessment technique to be used, and expectations about the assessment type, respectively, as seen in Table 18.

Table 18.

Institutional Expectations of Assessment

Code	f	Code Description
Assessment Technique	5	This code refers to institutional assessment expectations regarding what kind(s) of assessment technique(s) to be used by the teachers in their assessment practices.
Assessment Time	4	This code refers to institutional expectations in terms of when to assess students.
Assessment Type	4	This code refers to institutional assessment expectations regarding which type of assessment should be adopted while assessing students.

When examined, teachers often emphasise the assessment time while talking about institutional expectations. Their institution asks them to assess students at the beginning of the year, in the middle of the year and at the end. To illustrate, teachers were required to assess students at the beginning of the year to explore their English competency levels and put them into a suitable class. Also, teachers reported being asked to evaluate their students in the middle and at the end of the semester to collect information about their learning outcomes, grade them and decide whether students should pass the preparation class. Based on the emerging results, the end of a semester is found to be a reasonable time for assessment, and the institution asks teachers to assess their students at that time. The following excerpt sheds more light on the situation:

*My institution expects me to administer exams called mid-term in the middle of the semester and final exams **at the end of the semester**. Then we are expected to evaluate the students according to grades taken by students from those exams. (Pre-semi-structured interview, P3).*

Secondly, participants mentioned the assessment technique required by their institution. Almost all of them expressed that their institution expected them to use written exams to assess their students periodically. Based on the participants' statements, we can say that their

institutions understand assessment that focuses on results rather than development. The following extracts exemplify the situation:

*University expects us to **administer written exams** and give **mid-term and final grades** to students. (Pre-semi-structured interview, P1).*

*I'm an integrated class teacher. We teach speaking and pronunciation in the integrated course, but **the school tells me to assess students using the written exam technique**. The written paper will serve as an official document. I can't think of any other reason. (Pre-semi-structured interview, P2)*

Next, participants also talked about the expectations of their institutions regarding assessment type. They declared that their institution wanted them to assess their students in the middle and at the end of the semester and decide on students' success according to results derived from those practices. In addition, the participants stated that their institution did not let them behave freely regarding assessment and evaluation and that it did not allow teachers to apply any extra assessment in addition to the institution's planned assessment practices. For this reason, it was found that teachers had to assess their students at the end of the process and evaluate them at the end of the year with the data they obtained. Based on these emerging results, it can be indicated that teachers are compelled by their institutions to apply summative assessment. The excerpt below illuminates the situation better:

*Our institution requires us to **assess our students at the end of the semester and give grades accordingly**. **Considering these results, we determine the final grade, which will later be used for decisions on students' success** (Pre-semi-structured interview, P2).*

4.2.4.3. Influence of institution on Teachers' Assessment Perception and Practices

This emerging sub-theme provides information about the impact of institutional expectations regarding assessment on teachers' perception and practice of assessment. The participants listed the effects of the institution they work for on them. First of all, some of the participants stated that the institution did not have any impact on their assessment perceptions and practices. In addition, some of the participating teachers noted that the institution was also effective in their assessment frequency and the preference of the assessment technique they plan to use. Table 19 summarises the sub-theme:

Table 19.

Influence of Institution on Teachers' Assessment Perception and Practices

Code	f	Code Description
Influence on the assessment technique	7	Influence of institution on teachers' assessment technique. Change in teachers' assessment technique due to institutional requirements.
Influence on the frequency of assessment	6	Self-perceived influence of institution on teachers' beliefs and practices of assessment frequency. From more to less
No effect	3	Institutional expectations and requirements do not affect the teachers' knowledge, beliefs and practices.

Participants were asked to explain the influences of their institution on their assessment perceptions and practices, if there were any. First of all, teachers reported that the institution didn't influence their assessment perception at all. To illustrate, most participants noted that their assessment perceptions were already in line with how their institution perceived assessment. This result indicates that the participants' opinions about assessment which are in line with their institutions' did not cause any change in their perceptions of assessment. The following excerpts exemplify the situation:

Expectations of my institution didn't influence me at all. My beliefs and perceptions were similar. So, the institution didn't lead to any changes. (Pre-semi-structured interview, P2).

The institution that I work in didn't have any influence on my assessment perception. We already had similar conceptions. It didn't expect me to do something which contradicted me." (Pre-semi-structured interview, P3).

On the other hand, as a result of the analysis, some institutional influences on teachers' assessment perception and practices were revealed. Firstly, participants declared that their institutions affected their assessment. This result indicates that the participants' opinions about assessment in line with the institution did not cause any change in their perceptions of assessment. To illustrate, teachers who initially believed that students should frequently be assessed as much as possibly started to evaluate their students less often than they normally do due to institutional constraints. That is, they lessened the frequency of their assessment and reported that their practices influenced their ideas on the frequency of assessment. When prompted to explain their thoughts about the influence of the institution, P4 said:

I use sit-down written exams for my assessment practices. I didn't contradict my institution in terms of assessment. However, our institution asked me to decrease the

frequency of assessment, so I decreased it. My institution shaped my assessment practices in terms of frequency. I used to assess my students more frequently than I do now. Indeed, sometime later, I realised that it is useful and more practical. Instead of spending time preparing and grading exams, I focus on teaching. (Pre-semi-structured interview, P4).

In addition, it was found that the choice of assessment technique for the participants' applications was also affected by the institution they worked for. Teachers, who stated that they planned to prefer contemporary assessment techniques such as portfolios when they first started working at the institution, said they chose written exams due to the institution's limitations. Again, just like the change in their assessment frequency over time, it was revealed that teachers both frequently used written exams in their practices and thought that the ideal assessment technique was the written exam. The following excerpt shows how the institution influenced one of the participants' assessment technique preferences:

At the beginning of my teaching career, I used to think that assessment should be specific to individual students, like learning, and assessment should be planned by considering learner differences. In time, I started to be dependent on institutional expectations. I administer written and oral exams to all my students. However, I used to say I would implement portfolios or self-assessments in my assessment practices. I could not implement them. I just utilised written exams. They are also advantageous both for teachers and students (Pre-semi-structured interview, P1).

Taken together, these results suggest an association between institutional expectations and requirements regarding assessment preference, the ideal time for the assessment and frequency of the evaluation and teachers' perceptions and practices concerning these issues. Based on this, we can say that institutional expectations shape teachers' perceptions and assessment practices over time.

4.2.5. Teachers' Thoughts on the Influence of Educational Policy of the Country on Assessment

The last theme that emerged to explain teachers' perceptions of assessment is the assessment and educational policy of the country. This emerging theme presents research results on the issue of the assessment policy of the country and teachers' assessment perceptions and practices. Assessment and educational policy of country theme included two sub-themes: expectations of educational policy from teachers and influence of educational

policy on teachers' assessment perceptions and practices. Each sub-theme will be elaborated in the following sections.

4.2.5.1. Expectations of Assessment Policy from Teachers

This sub-theme gives information about the expectations of the country's assessment policy from teachers concerning assessment. From the answers given during the interviews, assessment-related expectations of the educational policy were found as expectations from teachers concerning time for assessment, expectations from teachers about assessment technique preference and expectations from teachers regarding assessment type. Each code is explicated in Table 20 and detailed in the following paragraphs.

Table 20.

Expectations of Assessment Policy from Teachers

Code	f	Code Description
Expectations concerning assessment time	8	Teachers' views on assessment-related expectations of the educational policy of the country from teachers regarding assessment time – end the end of some units and semester.
Expectations concerning assessment technique preference	8	Teachers' views on assessment-related expectations of the country's educational policy from teachers regarding assessment technique preference- at the end of some units and semesters.
Expectations concerning assessment type	4	Teachers' views on assessment-related expectations of the educational policy of the country from teachers regarding assessment type – diagnostic and summative assessment.

It can be seen that teachers were required to implement written exams in their assessment practices. In other words, when we examined how the teachers express what is expected concerning assessment techniques from them by the assessment policy of the country, we found a similar pattern in all explanations. All the teachers reported being required to implement written exams in their assessment practices. Based on the results, a visible consistency among teachers' assessment perception, institutional assessment-related expectations and assessment-related requirements set by the assessment policy of the country can be seen. In addition, based on the results, we can say that the assessment technique preference of the teachers is shaped, namely determined, both by the country's education system and by the institution's expectations. One of the participants exemplifies the situation:

Our educational policy requires us to assess our students two or three times a semester using written exams. Then it lets us decide whether students should pass or not. When I was working for MONE, namely when I was teaching English in a state

*secondary school, I was also **required to implement written exams** and evaluate students according to the results I derived. (Pre-semi-structured interview, P3)*

In addition, participants noted that there were certain times when they were required to assess their students. They added that the time for assessment was determined by the educational policy as well. Overlapping words, which are presented in bold below, demonstrate the consistency among teachers' explanations concerning policy-related requirements that they were exposed to. It was understood from the emerging results that teachers are not free to specify assessment time for their practices. They are also required to assess their students three times in different time intervals throughout a semester. The explanations below depict the situation:

*The **policy** wants us to **assess our students at certain periods** through 1st, the 2nd, and 3rd, exams or mid-term or final exams and **decide on students' success at the end of the semester**. (Pre-semi-structured interview, P3).*

*The country's assessment policy also shapes the school's expectations. When I was working in national education, we wrote only grammar and reading questions. Speaking was ignored. I am currently teaching English at university. The situation is still the same. Thus, we are obliged to **use a written exam**. (Pre-semi-structured interview, P2).*

Finally, the teachers stated that the assessment practices only focus on the result and mostly consider the final exam results at the end of the year when evaluating the students. Similarly, teachers must adopt a summative assessment type in their actual assessment practices. It was revealed as a result of the analysis that, according to teachers, this grading system was not determined by themselves. Instead, this assessment system deciding on success by looking at the end result was caused by the country's education policy.

4.2.5.2. Influence of Educational Policy on Teachers' Assessment Perceptions and Practices

Finally, this sub-theme includes research results that refer to the impact(s) of the country's assessment policy on teachers' assessment perceptions and practices. According to the findings that emerged as a result analysis, the teachers were found to be influenced by the assessment policy of the country regarding the frequency of assessment practices and assessment technique(s) to be used. In addition, one participant stated that the country's assessment policy has a restrictive effect on him. Table 21 presents the sub-theme:

Table 21.

Influence of Educational Policy on Teachers' Assessment Perceptions and Practices

Code	f	Code Description
Influence of the assessment technique selection	9	Influence of assessment policy of the country on teachers-technique selection – adopting traditional ones especially written exams.
Impact on the frequency of assessment	4	Influence of assessment policy of the country on teachers-frequency of implementation of assessment – from more frequently to less.
Restrictive influence	3	Influence of assessment policy of the country on teachers-restrictive limited role.

First of all, participants proposed a change, namely an influence, in their assessment perceptions and practices due to the influence of the country's educational policy. To elaborate, one of the participants indicated an influence of educational policy on his/her frequency of assessment. S/he claimed that contrary to his/her initial perceptions, the country's educational policy led him/her to think that students may not be assessed so frequently. Instead, teachers should put more emphasis on teaching. Based on this, the teacher was found to believe and practice according to his/her changing cognitions regarding the frequency of assessment.

*I worked at MONE for about ten years. In English curricula, **exams were administered only three times per semester**. Over time, I started to think that teaching the subjects in the curriculum should be emphasised **rather than assessing and evaluating students too frequently, and I applied it this way**. (Pre-semi-structured interview, P2).*

Furthermore, an influence on teacher assessment practices regarding technique preference was identified as a result of the analysis. They claimed that thanks to the influence of the country's educational policy, they do not have any right to choose a suitable assessment technique for their practices. To sum up, teachers' assessment techniques preference was found to be determined by the educational policy of the country. The following excerpt illustrates the situation:

I have to use written exams. I think so. I don't have any other choice. I have to obey these rules. The educational policy of the country influenced my assessment practices. (Pre-semi-structured interview).

Moreover, one of the participants highlighted that educational policy considering assessment limited their practices regarding, for example, assessment technique preference.

Due to those requirements, s/he reported that s/he didn't put what s/he had in their mind considering assessment.

I want to use a slightly more contemporary assessment. For example, I would like to use portfolios and peer assessments, **but we are not allowed to apply them officially.** We do not **have a chance to evaluate students** with these assessment techniques. So this **limits** us. This was also the case when we were working at MONE. **The ministry determined everything.** (Pre-semi-structured interview, P2)

These results indicate that the country's assessment policy causes teachers to adopt an assessment approach they do not believe in. In addition, emerging findings showed that although teachers initially did not appreciate the expectations of education policy from themselves regarding assessment, over time, they began to accept them, and they thought and acted in line with the assessment policy of the country.

4.3. What are the Participants' Strengths and Weaknesses Considering Language Assessment as They Perceive?

4.3.1. Teachers' Strengths Related to Assessment

Within the scope of the study, the strengths of the instructors regarding language assessment and evaluation were also tried to be revealed. For this, a pre-semi-structured interview was applied to all participants. In addition, teachers were observed in their classrooms to see whether their claims match or mismatch to their classroom practices, and a follow-up semi-structured interview was arranged for each.

Table 22.

Themes and sub-themes about strengths of the teachers concerning language assessment

Theme	Categories
Teachers' strengths concerning assessment results.	Teachers' other strengths in assessment results Teachers' strengths in assessment result usage Teachers' strengths in assessment data evaluation
Teachers' assessment techniques related strengths	Teachers' strengths in assessment technique knowledge and implementation Teachers' strengths in considerations for selection of techniques Teachers' strengths in purpose for assessment technique use
Teachers' strengths in developing and administering exams	Teachers' strengths in test preparation Teachers' strengths in administering exams Teachers' test item related strengths

Teachers' strengths about introductory topics in assessment	Teachers' strengths in assessment related definitions Teachers' strengths in purpose of assessment
Teachers' strengths in ethical considerations in assessment,	Teachers' strengths regarding views of ethics Teachers' strengths in ethical practices
Teachers' strengths in scoring and grading	Teachers' strengths in grading Teachers' strengths in scoring
Teachers' other strengths regarding assessment	Teachers' other assessment related strengths Teachers' strengths in avoiding from measurement error

As a result of the analysis, seven broad themes were identified, indicating participants' self-perceived strengths in assessment. They are teachers' strengths concerning assessment results, teachers' assessment techniques-related strengths, teachers' strengths in developing and administering exams, teachers' strengths about introductory topics in assessment, teachers' strengths in ethical considerations in assessment, teachers' strengths in scoring and grading and teachers' other strengths regarding assessment. In the following sections, emerging themes will be presented in detail with excerpts from the transcripts.

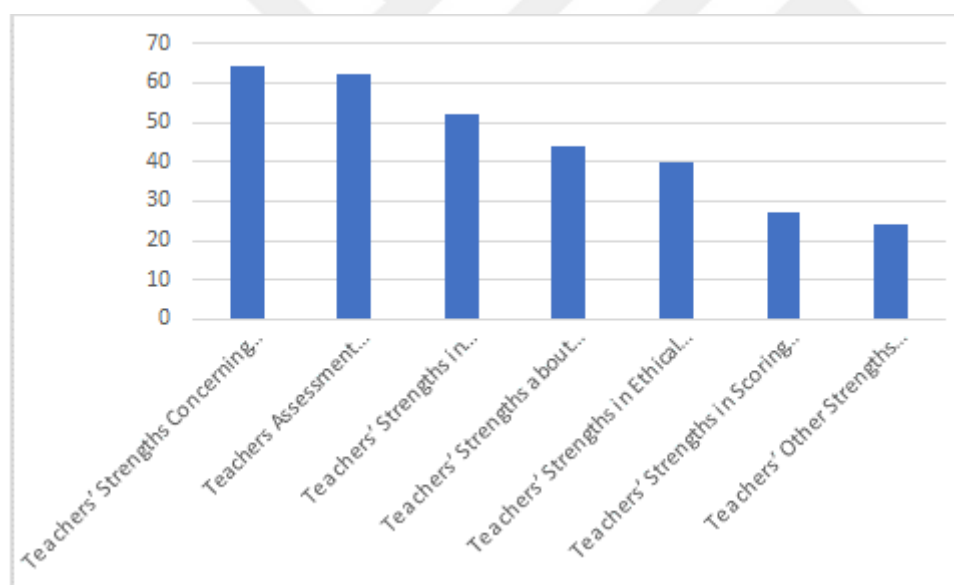


Figure 4. *Teachers' Strengths Related to Assessment*

Before teachers attend RRD sessions, their strengths in assessment, especially language assessment, are revealed. Based on their statements, the areas in which teachers perceive themselves as strong are listed according to frequency of citation. Based on the analyses, the main area in which teachers perceive themselves as strong are their strengths related to assessment results. In addition, among the seven themes that emerged, the area in which the teachers perceived themselves as the strongest with the lowest frequency of citation was found to be other strengths in assessment, including their measurement error. In addition,

teachers' perceived strengths in scoring and grading can be said to be the area they feel least powerful since it has nearly the lowest frequency of citation. The following sections show teachers' strengths in language assessment.

4.3.1.1. Teachers' Strengths Concerning Assessment Results

In this section, teachers' perceived strengths regarding assessment results showed itself as a prominent theme. The participating instructors often emphasised assessment results as an important element in their assessment practices. The sub-themes that emerged in the light of the analysis results are teachers' other strengths in assessment results, teachers' strengths in assessment results usage, and teachers' strengths in assessment data evaluation. Detailed explanations belonging to each sub-theme are presented in the following parts below.

4.3.1.1.1. Teachers' Other Strengths in Assessment Results

With this sub-theme, the perceived strengths of the participants related to assessment results were revealed. These emerging self-perceived strengths are related to interpretation of the assessment results, reflection of the assessment results by both students and teachers, sharing the assessment results with stakeholders and practices performed to obtain reliable assessment results. In the following parts, detailed explanations about the sub-theme are provided by presenting the codes constituting the sub-theme, frequencies of codes, and code descriptions are given as they are also seen in Table 23 below.

Table 23.

Teachers' Other Strengths in Assessment Results

Code	f	Code Description
Interpreting results regarding teacher efficiency	6	Participants explain that they make some interpretations about assessment results considering their teaching and its effectiveness.
Way of reflection of assessment results by students	6	Participants explain how should reflection of assessment results be expressed by the students.
Way of reflection of assessment results by teachers	5	Participants explain how should reflection of assessment results be expressed by the teachers.
Use of web pages in sharing assessment results	4	With this code, participants show how they inform stakeholders about assessment results.
Being objective for producing reliable assessment results	3	Participants believe that it is necessary to be objective in order to obtain reliable data.
Being careful for yielding reliable assessment results	2	Being alert is an important factor for getting reliable assessment data. Participants believe that teachers should be careful to achieve this.

It is important to note that participants perceive assessment results as crucial for them basing their opinion on the ground that they are teachers and their ultimate aim is to lead students to have better learning experiences. Considering this, they report interpreting assessment results in order to evaluate their teaching efficacy namely, to find out the weaknesses in their teaching and correct them accordingly. In other words, teachers interpret results to see whether they are effective for their students' learnings as a teacher or not and to increase the quality of the instruction which in turn raises students' learning. The following excerpt illustrates the situation:

*Assessment results are important to me. From my point of view, I can say that by looking at the assessment results, I ask myself these questions **Did students succeed or not, and where am I in this situation? If the student succeeded, how much could I help, or if the student failed, why couldn't I help?**" This is how I interpret the assessment results. As a teacher, I interpret the results to determine how effective I have been or not. (Pre-semi-structured interviews, P4)*

Participants also touched upon how they thought reflectively on the assessment results and indicated that the way they administered the exam, the way through which they scored and graded, and the assessment technique/tool they employed while they were assessing their students were the main areas they usually reflected based on obtained assessment results. To elaborate further, the teachers stated that by looking at the data they got through their assessment practice, they would reason about what went well and what went wrong during the exam and be more careful in their next applications. In addition, the teachers expressed that they reflected on the effectiveness of the assessment technique/tool they used while evaluating their students according to their exam results. In other words, it was determined as a result of the analysis that the teachers reflected on whether they could get valid, reliable and error-free assessment data with the technique or tool they utilised. If there were any problems related to these issues, they made decisions to improve the quality of their following practices by trying to correct them. In short, they addressed making adaptations in their assessment practices if necessary.

Considering the assessment results, I say to myself "Look, I used the X technique/tool, I took individual differences into consideration when choosing the assessment tool or technique, I gave clear instructions to the students, I tried to evaluate each student individually, I was objective. For these reasons, the students' exam results were high. The students learned. On the contrary, my thoughts might be

the opposite. "I assessed the students without any research or without taking anything into consideration. For this reason, the students received low grades, I have to make changes." This is how the reflection of the assessment results should be for me. I mean, I am trying to achieve this as I mentioned. (Pre-semi-structured interview, P3)

From these statements, it might be deduced that the teachers were able to be comfortable with benefitting from assessment results which made it possible for them to improve their teaching quality.

4.3.1.1.2. Teachers' Strengths in Assessment Results Usage

EFL instructors highlighted that they are aware of the importance of assessment results. They were found to believe that these results shed light on their own and students' paths. That's why teachers were of the opinion that the assessment results can be employed to make decisions on students' success, to identify their weaknesses, to reveal the learning status of students, and to make adaptations in the assessment tool, teaching plans and even the curriculum when necessary. Table 24 shows codes constituting this emerging sub-theme.

Table 24.

Teachers' Strengths in Assessment Result Usage

Code	f	Code Description
Use of assessment results for identifying student learning	9	Participants mean that they use assessment results in order to see whether students have learnt the subjects taught to them or not.
Making changes in teaching by looking at assessment results	6	Participants perceive assessment results as important because s/he reports that these results are used while making some necessary changes in her
Using assessment results in identifying students' failures	2	Participants explain that they utilize assessment results in order to identify students' weaknesses in order to better help them.
Using assessment results in adapting assessment tool	1	Participants state s/he adapts the assessment tool that s/he uses by taking assessment results into consideration. Necessary adaptations in assessment tools are made based on derived results.
Strength in making adaptations in curriculum based on results	1	Participant reports that he / she changes the curriculum based on assessment result. When it is really necessary, s/he makes necessary changes in it.
Use of assessment result in deciding on student success	1	Participant explains that s/he utilizes assessment results in order to decide on whether students should pass the class or not.

As it is clear in Table 24, participants utilised assessment results for various purposes. First, teachers' primary usage of assessment results is to determine the learning levels of the students. In other words, it can be expressed that teachers determine whether students have learnt the subjects taught to them by their teachers by taking into account the results obtained as a result of assessment practices. In parallel with the results concerning teachers' self-perceived assessment purposes and perceived assessment technique usage, it can be said that the priority of teachers in assessment and evaluation is to determine the learning levels of students after teaching. The following excerpt explicates the situation:

In other words, I mostly use the assessment results to reveal whether students have understood the subjects and to see the student's progress. Has the student learned what he/she needs to learn or not? I can think about the development of the students by using the assessment results. (Pre-semi-structured interview, P1)

Teachers also reported using the assessment results in evaluating their instructions. On the basis of findings revealed so far, we can say that teachers' opinions about the use of assessment results were also related to increasing the quality of their teaching by determining their strengths and weaknesses and they were found to be in line with their perceived purpose of assessment and assessment technique used. Therefore, it might be claimed that teachers were clearly in favour of using assessment results to make necessary changes in their teaching plans and practices for the good of better teaching and learning practices. This notion was explicated below:

I make changes in my teaching using assessment results. For example, the written exam result tells me "The student did not succeed in unit 2". I make this inference using the results. I will go back to unit 2 in the next lesson. I say to students "Look, you couldn't do this in the exam, it would be like this," and I teach that topic to students again. Also, I change the materials I use in the course other than the course book and the activities we do in the lesson. I use the results for the purposes I mentioned. (Pre-semi-structured interview, P4)

4.3.1.1.3. Teachers' Strengths in Assessment Data Evaluation

Instructors, when presenting their opinions regarding the assessment results, also referred to the assessment data evaluation. Although it is not as prominent as the two previous categories, participants' perceived strengths in assessment data evaluation are still worth being classified under a sub-theme. From the explanations of the participants, it was revealed

that they evaluated the assessment data with the specific methods they devised without using any statistical technique. In light of these evaluations, the participants stated that they made decisions about the success of the students and find out missing and completed parts both in students' learning, their teaching and the curriculum. In other words, through investigation, the teachers perceived strengths related to the evaluation of the assessment data were found as determining students' learning, evaluating the effectiveness of teachers' teaching, and making adaptations on to the curriculum if it is necessary. Table 25 presents the details of this emerging sub-theme.

Table 25.

Teachers' Strengths in Assessment Data Evaluation

Code	f	Code Description
Evaluating assessment data for student learning	9	Participants report evaluating assessment data to improve student's learning situations, to achieve better learning outcomes.
Evaluation of assessment data in deciding on students' success	4	Participants declare that they evaluate assessment data in order to see students' levels and make a decision about their success.
Assessment data evaluation considering teaching	3	Participants believe that through assessment data evaluation, teachers evaluate the effectiveness of his /her teaching and identify his / her strengths and weaknesses considering teaching.
Evaluation assessment data linking to curriculum	2	Participants report that by analysing assessment data, they can have information related to the implementation of the curriculum.

To start with, the majority of the instructors mentioned that they evaluated the assessment data to support the development of students. They evaluated the assessment data in order to determine the learning level of the students and make comments about the development of the students and support them by identifying their needs and giving feedback accordingly. Based on this emerging finding, teachers can be said to implement formative assessment namely assess for learning. Their ideas were exemplified below:

*First of all, I would like to point out that the **teacher does not give the grade, the student receives it. In my opinion, assessment data should be evaluated by teachers in determining the learning level of students and revealing the development at this level throughout the process. I do so. (Pre-semi-structured interview, P3)***

*I utilise the data to **evaluate the student**. Sometimes this evaluation is about student achievement and I decide their pass/fail decisions. Sometimes this evaluation is to **identify the student's weaknesses and guide them for the better**. For this purpose, I assess my students and evaluate the data I have obtained. (Pre-semi-structured interview, P2)*

Secondly, it was revealed that the primary purpose of almost all of the participants in evaluating the assessment data was to make a decision about the success of the student. Since all the participants worked in the same institution and the assessment expectations of the institution along with the student profile addressed were the same, the participants stated that they first evaluated the assessment data to decide on whether the students should pass the class or not. Participant clearly exemplified why s/he evaluated assessment results:

*Assessment and evaluation data, that is, the data we obtain from the assessment results. Teachers can evaluate the data **when making comments about a student's level or making decisions about whether a student should pass or fail**. They can be evaluated to say "the student passed the class". That's how I apply it in my practices. (Pre-semi-structured interview, P3).*

In a nutshell, participants indicated that they used and evaluated assessment results in their actual classroom practices. For this fact, they always foregrounded that results are important in terms of turning their teaching and students' learning into more effective processes.

4.3.1.2. Teachers' Assessment Techniques Related Strengths

In this case study, the participants indicated their strengths related to assessment techniques. Sub-themes included in this theme are teachers' strengths in assessment technique knowledge and implementation, teachers' strengths in considerations for selection of technique, teachers' strengths in purpose for assessment technique use. Each sub-theme is explained in detail in the following parts.

4.3.1.2.1. Teachers' Strengths in Assessment Technique Knowledge and Implementation

Having the highest frequency, known and used assessment technique was found as a very prominent sub-theme which was central to participants' strengths related to assessment techniques. With this sub-theme, it can be said that the techniques that most of the participating instructors know and apply in the classroom are mostly traditional techniques and they do not know and cannot implement any of alternative assessment techniques in their actual assessment practices. Table 26 shows details regarding the sub-theme.

Table 26.

Teachers' Strengths in Assessment Technique Knowledge and Implementation

Code	f	Code Description
Practicing sit-down written exams	11	Self-perceived strength in practicing sit-down exams. Participants claim that they are good at administering sit down paper and pencil test as an assessment technique.
Knowing traditional assessment techniques	8	Participants see themselves competent about knowledge regarding traditional assessment techniques.
Practicing traditional assessment techniques	8	Self-perceived strength in practice of assessment method focusing on only traditional ones. Sit-down paper and pencil exams, oral exams, multiple, multiple-choice exams.
Practicing performance assessment	2	Self-perceived strength in assessment method used in classroom applications – performance assessment.

As can be seen from Table 26 above, the teachers know and mostly use written exams in their practices. The reason behind this was also expressed by the teachers and revealed the possibility that teachers' previous experiences when they were students influenced their knowledge and practice of assessment techniques. Teachers reported being assessed through written exams in their English courses not only in their high school but also in their undergraduate years. Even though they were taking assessment courses and learning alternative techniques, they were again assessed with written exams. This experience led them to think that written exams are ideal ways to measure students' competency levels. Most probably, they might have begun to know and implement written exams. Based on the analysis, we can infer that the strength of the teachers regarding assessment techniques was limited to only classical assessment techniques. The following excerpts from the data exemplify this situation:

*I use almost **only written exams** and sometimes **oral exams** in my assessment practices. I can say that this technique, **applying the written exams, is my strength.** (Pre-semi-structured interview, P1)*

*There are various assessment techniques, but I **consider myself strong in using traditional ones such as written exams** I mostly know and apply traditional assessment techniques. (Pre-semi-structured interview, P4)*

*When I was in high school and at university, I was **assessed and evaluated through written exams.** Even in the course where we learned alternative assessment techniques in the faculty of education, **written exams were used.** The school also asks*

me to administer the written exams. Therefore, I do not need to learn and apply an extra technique. (Pre-semi-structured interview, P2)

Next, some of the participants expressed in the interviews that they employ performance assessment in their practices. To give detailed information about the subject, it was found that teachers only associate performance assessment with speaking skills. In addition, although teachers claimed that they applied performance assessment during a classroom observation, they did not use any rubrics and did not enable their students to use English through authentic tasks and activities. As a result of the analysis of the interviews and the classroom observation, it can be concluded that the participants partially had strengths in the performance assessment, but needed further improvement about this issue. The following excerpt illustrates this finding:

Although the participant stated that he knew and applied performance assessment in the interview, his application was not observed in the observations. s/he just made students speak in English. In the post-interview, performance assessment questions were asked again, and it was concluded that the participant had knowledge about this technique, but this information was incomplete. (Researcher reflection report)

4.3.1.2.2. Teachers' Strengths in Considerations for Selection of Techniques

This sub-theme presents findings concerning participants' perceived strengths related to what they consider when choosing an assessment technique for their practices. Based on the analysis, teachers considered content validity and practicality as priorities in choosing a technique. In addition to these, teachers' knowledge, and reliability were also considered factors in the selection of assessment techniques. Table 27 presents teachers' considerations for the selection of assessment techniques.

Table 27.

Teachers' Strengths in Considerations for Selection of Techniques

Code	f	Code Description
Content validity as a criterion for selection	11	Self-perceived strength in choosing assessment technique – content validity is the most frequently uttered criterion.
Practicality as a criterion for selection	9	Perceived strength in choosing assessment technique. Practicality is considered important. Techniques should be practical in administering and grading.
Considering student level in the election	3	The perceived strength of participants about assessment technique selection. They are good at choosing assessment technique/tool which is in line with students' competency levels.
Ensuring construct validity	2	Self-perceived strength in choosing assessment technique – construct validity- participants are of the opinion that they should be careful about the fact that the thing they assess should match the thing they actually assess in their assessment practices.
Reliability as a criterion in the selection	1	Self-perceived strength in choosing assessment technique – reliability- The participant knows how to choose an assessment technique which is in line with the objectives and should help him/her assess the students without measurement error. Results obtained from this tool/technique should reflect reality.
Considering teacher knowledge	1	Self-perceived strength in choosing assessment technique- teacher knowledge. It is believed that teachers should know how to implement the tool or assessment technique. Considering this, s/he chooses the assessment tool/technique for his/her practices.

Participants were asked to explain their thoughts on what they considered while selecting an assessment tool/technique for their practices. Firstly, they talked about the factors they considered when choosing written exams and declared that they first considered the content of the exam. In other words, they were careful not to ask questions beyond the scope of the subjects that they taught. All of the teachers reported only choosing an assessment technique/tool which totally assesses students depending on previously taught subjects. The following excerpts exemplify this situation:

*When I asked some questions to her about her selection of any assessment technique or assessment tool. She mentioned that assessment tool or method **should be in line with what I taught to my students and it should be easy to administer and score.** Then she did not mention extra things. (Researcher reflection)*

*When **choosing an assessment technique or tool**, the only thing I consider is **whether it is parallel to the subjects I have covered or not.** (Pre-semi-structured interview, P1).*

Moreover, the participants mentioned that they prioritized practicality when choosing an assessment technique. For them, the technique should be easy to administer, easy to grade, and should not cost a lot of time and money. From the teachers' own terms, we can deduce that teachers did not want to face problems while they were assessing their students. Also, they may want to assess their students without spending too much time and effort.

*I would make sure that the technique or tool I would choose should be **practically applicable to the student and the teacher** It should not **necessitate too much effort and time from me** because we have lots of responsibilities. (Pre-semi-structured interview, P3).*

In addition to these, participant teachers stated that they also considered reliability when choosing any technique or tool for their assessment practices. From the explanations given by the teachers, we can understand that they associated reliability with preventing cheating. That's why we found that teachers preferred assessment techniques which were well-developed and did not have any missing parts which may create opportunities for students to cheat. The following bold statements explicate the situation:

*The technique I will choose **should not allow for cheating**. In this way, **I can measure the information in the minds of the students**. I can **easily ensure reliability** with a well-organized written exam. (Pre-semi-structured interview, P3)*

These findings indicate that teachers do not choose the technique that they will employ randomly. They took practicality, construct validity and reliability into consideration. By considering these, they thought that they could assess and evaluate their students effectively.

4.3.1.2.3. Teachers' Strengths in Purpose for Assessment Technique Use

This sub-theme provides findings referring to teachers' perceived strengths concerning the purpose of implementing assessment techniques. As a result of the analysis, we found that teachers used assessment techniques to find out the students' learning outcomes and explore the quality of their instruction. Table 28 presents the codes included in this sub-theme.

Table 28.

Teachers' Strengths in Purpose for Assessment Technique Use

Code	f	Code Description
Using technique for identifying students' competency level	4	Participants employ assessment technique that they choose in order to reveal students' level of learning. This code is related to self-perceived strengths of participants considering employment of assessment technique.
Using techniques for identifying teaching effectiveness	2	This code is related to self-perceived strengths of participants considering employment of assessment technique. They report to use them for evaluating their effectiveness in teaching.

As can be seen from Table 28, by using assessment techniques, teachers first aim to shed light on whether their students could learn the subject or not. Actually, phrases 'identifying students' learning levels, checking their understanding, evaluating teaching efficacy' come to the fore in almost all of teachers' answers. Based on these results, we can say that teachers' perceived purposes of implementation of assessment techniques and their previously emerged self-perceived assessment purposes, and their perceived definition of assessment were found to be in line with each other. These common points in teachers' answers demonstrate a visible consistency. Considering these, these emerging findings can be interpreted as showing a certain logical pattern by confirming and supporting one another.

4.3.1.3. Teachers' Strengths in Developing and Administering Exams

Teachers within this study also stated their strengths in developing tests and administering them. These perceived strengths were classified into three sub-themes such as teachers' strengths in test preparation, teachers' strengths in administering exams, teachers' test item related strengths. Each sub-theme is explained in detail with the help of tables.

4.3.1.3.1. Teachers' Strengths in Test Preparation

This sub-theme is the prominent sub-theme of the related theme. Under it, the instructors' knowledge about test preparation, namely their perceived strengths, are brought together. To make this statement clearer, participants always prioritize content validity while preparing a test. It is also seen as important by the teachers that the test should be practical for both students and instructors. In addition, it has been revealed that the participants see themselves as strong in preparing written exams and not including questions whose answers might be predicted by the students so as to ensure the reliability of the exams. Table 29 demonstrates the sub-theme.

Table 29.

Teachers' Strengths in Test Preparation

Code	f	Code Description
Ensuring content validity when preparing a test	21	Perceived strength in test preparation. As for validity, content validity is only regarded.
Considering practicality while preparing tests	3	Perceived strength in test preparation. Practicality for both teachers and students is regarded as important and it is taken into consideration during test preparation.
Strength in preparing written exams	3	When asked about assessment technique, sit down exam was shown as the most frequently used and known techniques. Strength in preparation of written exams.
Preventing guessing for ensuring reliability	1	Perceived strength in test preparation through preventing guessing for enhancing reliability.

Table 29 illustrates teachers' perceived strengths concerning test construction. It is apparent from this table that nearly all of the participants firstly consider content validity while constructing their tests. That is, the teachers indicated that while they were preparing a test, they only asked questions based on the subjects that they taught to the students, and they did not go beyond these subjects while preparing the questions. Participant 1 stated how s/he ensured content validity while preparing the tests:

While preparing the test, I always have a list of the subjects that I taught to the students in front of me and I come to the beginning of the list mainly from the last subjects we covered. I ask questions about the subjects I have taught. I do not ask questions about subjects that I have not covered. (Pre semi-structured interview, P1).

In addition to content validity, EFL instructors also considered the practicality of the test while preparing it. To them, exams should be practical for both students and teachers. As it emerged previously, teachers wanted their exams be practical in terms of administering, scoring and grading. Teachers showed preparing practical tests as one of their strengths in pre- and post-interviews, although they could not be observed doing so in classroom observations. Participant 2 stated how s/he prepared practical tests in addition to content validity.

The exams I prepare should be easy to apply to the student. Then it should be easy to grade. I consider myself competent in preparing exams with these features. (Pre-semi-structured interview, P2).

It was determined that teachers felt strong themselves especially in preparing and applying written exams. Especially when preparing the written exams, ensuring content validity and practicality was revealed as teachers' perceived strengths with regard to test construction.

4.3.1.3.2. Teachers' Strengths in Administering Exams

Another sub-theme that emerged as a result of the content analysis is related to the teachers' perceived strengths in administering the exams. All of the participants showed more than once that their most distinctive strength on this topic was their efficacy in preventing students from cheating during exams. In addition, the codes such as the physical condition of the classroom, giving clear instructions to the students, and the sitting arrangement of the students during the exam were identified as their perceived strengths. Table 30 explains the relevant sub-theme in detail.

Table 30.

Teachers' Strengths in Administering Exams

Code	f	Code Description
Preventing cheating during exams for reliability	12	Self-perceived strengths in administering exams. Most frequently uttered point to be considered during exams is preventing cheating for enhancing reliability.
Organizing physical condition of class	2	Self-perceived strength in administering exam which is related to physical condition of class in which exams are held.
Giving clear instructions in administering exams	1	Self-perceived strength in administering exam which is related to comprehensibility of instructions given during the exam.
Suitability of students for the exam	1	Self-perceived strength in administering exam which is related to preparing students physically and psychologically for assessment before the exam.
Way of seating	1	Self-perceived strength in administering exam which is way of seating. Here, it is perceived that students' way of seating in class before the exams is important.

As can be seen from Table 30, teachers' the first and foremost perceived strength in administering exams is preventing cheating. They intended to enhance reliability during the exams through preventing cheating. Whenever teachers were asked to express their views concerning application of exams, they immediately mentioned cheating. Taken together, this finding suggests that there is an association between teachers' ideas on administering exams and preventing cheating. Also, they reported increasing reliability during exams through preventing cheating. That's why teachers stated several times during the interviews that they absolutely did not allow cheating during exams. The excerpt below exemplifies the situation:

*In fact, **the way I administer exams** in general is as follows: I stand by the students during the exam, **I prevent the students' cheating**. This is also **how I ensure reliability**. Apart from these, **I don't know what to pay attention to in the exam application** or how to increase the reliability in this process. (Pre-semi-structured interview, P3).*

In addition, the teachers reported being sure about physical appropriateness of the classes before administering exams. According to them, the heat, light and fresh air conditions of the classroom influence how students perform during the exams. In order to avoid those negative effects, the teachers stated in the interviews that they checked suitability of the classes in this respect both before and after the exam. Participant 2 stated how s/he prepared the classrooms before administering the tests:

*What should I pay attention to while applying the exams? **I make sure that the physical environment of the classroom is suitable for the students**. This convenience can be **related to sound, light, enough space**... I pay attention to such things while the students are answering the questions in the exam. (Pre semi-structured interview, P2).*

Other codes belonging to this sub-theme and detailed explanations about them can be found in the Table 30 above. As a conclusion, we can say that the strengths of the participants about the exam applications are for the students. The strengths mentioned in preparing sitting arrangement of the classroom for the exam and ensuring that the classroom is large and bright are aimed by the teachers to prevent students from cheating. Just as teachers' perceptions of assessment focus only on administering exams, teachers' strengths about administering exams revolve around preventing students from cheating.

4.3.1.3.3. Teachers' Test Item Related Strengths

In this sub-theme, findings related to teachers' perceived strengths in what to consider while including test items in a test. In addition, the participants' perceived strengths regarding their purposes while choosing the test items for the test. Table 31 gives detailed information about the results:

Table 31.

Teachers' Test Item Related Strengths

Code	f	Code Description
Selection of test item for identifying students' level of learning	2	Perceived strength in purpose for choosing test item and including it in a test. Here participants include test items in a test in order to discover students' competency levels.
Selection of test item for distinguishing hardworking from poor one	2	Perceived strength in purpose of test item and including it in a test. Distinguishing students.
Enhancing variability in including test items in a test	2	Being cautious to include different kinds of test items in order to ensure variability. Teachers see themselves as competent in creating variability in a test in terms of test items.
Selection of items for a test based on students' likes	1	Perceived strength related to test item. Taking students' likes, interests into consideration during test preparation. Choosing or writing test items depending on students' likes is considered to be a strength by the participant.

During the interviews, teachers first talked about their goals while choosing test items for the test they prepared. These statements were found to be similar to the previous findings of the investigation. To be clearer, while adding test items to the tests, the participants cared choosing test items that aim to find out whether the students learnt the subject taught to them or not. In other words, the participants aimed to determine the learning levels of the students while choosing the items for the test. This result is also parallel with teachers' perceived assessment purposes. Both of them aim to shed light on students' learning gains. The following excerpt exemplifies the situation:

Why do I include test items in a test? To see whether the students learn the subjects taught to them or not? I choose or write appropriate test items according to my purpose and then include them into the test. (Pre semi-structured interview, P3).

Later, teachers' perceived strengths concerning what they consider while selecting test items for a test were revealed. Based on the findings, it can be said that when choosing test

items and including them in a test, the participants chose a test item that would be useful in distinguishing those who really knew from those who did not know, as shown in the following excerpt:

The test items that I choose or write must distinguish students, that is, they must distinguish between those who know and those who do not. I prepare test items or choose them accordingly. Sometimes I prepare difficult tests and include difficult questions. Then I use those questions, as I said before, to distinguish those who really know from those who don't. I don't have any other purposes. (Pre-semi-structured interview, P2).

4.3.1.4. Teachers Strengths About Introductory Topics in Assessment

This emerging theme presents participants' perceived strengths related to basic introductory issues in assessment. These are teachers' strengths in assessment-related definitions and teachers' strengths in purpose of assessment. Both of the sub-themes have been explained in the following parts.

4.3.1.4.1. Teachers' Strengths in Assessment Related Definitions

With this sub-theme, the teachers' perceived strengths about basic definitions related to assessment are illustrated. In other words, teachers' knowledge about definitions of assessment, measurement error and reliability are provided. Table 32 illustrates the results below.

Table 32.

Teachers' Strengths in Assessment Related Definitions

Code	f	Code Description
Definition of measurement error	3	Perceived definition of measurement error
Definition of reliability	1	Perceived definition of valid assessment result
Definition of valid assessment result	2	Perceived definition of valid assessment result
Definition assessment regarding diagnosing student's current level	2	Perceived definition of assessment
Assessment as a process of determining achievement of objective	2	Perceived definition of assessment

Definition of assessment through tracing student's development	2	Perceived definition of assessment
Definition of formal vs informal assessment	4	Perceived definition of formal vs informal assessment
Definition of correct evaluation	3	Perceived definition of correct evaluation
Definition of incorrect evaluation	4	Perceived definition of incorrect evaluation
Definition of formative and summative assessment	1	Perceived definition of summative vs formative assessment

As a result of the interviews, it was concluded that some of the participants knew some definitions while they also had incomplete knowledge about some of them. When we look at the Table 32 in light of these explanations, the basic definitions concerning assessment were somehow known by all or some of the participants, but when their answers are examined, it has been revealed that they have also weaknesses in these definitions. For this reason, it also be noted that teachers' correct answers were regarded as their strengths. Teachers' incomplete or wrong explanations were accepted as their weaknesses and they will be mentioned again in the weaknesses section of the research.

While defining the assessment, the participants focused on three different parts and defined it separately. These are to follow the student progress throughout the process, to determine whether the students reach the goals at the end of the process, and to reveal the existing knowledge of the students. From their answers, we can the assessment types that teachers adopted in their practices. Their definition of assessment focusing on tracing students' progress shows their formative purpose, while their definitions of assessment focusing on identifying students prior knowledge before the instruction signifies their diagnostic purpose. The excerpts below show teachers' perceived strengths focusing on teachers' various perspectives regarding assessment definition.

*There are **target behaviours** that we expect from students. It is **an assessment** to **determine** whether **they have these target behaviours** or the level of knowledge we expect from the student at work. (Pre-semi structured interview, P3).*

*In other words, as I mentioned before, assessment is **the process of informing students** and us about their progress from beginning to the end. We can identify and follow students' **development** through assessment practices. (Pre-semi structured interview, P1).*

*In the interview, the participant explained that assessment is a practice through which teachers determine students pre-knowledge at the beginning of a unit, course or lesson. During the classroom observation, I observed that teacher asked some questions before the lesson. When I asked her/him the reason of it, s/he replied that s/he tried to discover what students know. It can be deduced that assessment is a process in which teachers **discover what students already have**. (Researcher reflection report)*

4.3.1.4.2. Teachers' Strengths in Purpose of Assessment

This sub-theory gives information about teachers' perceived strengths considering purpose of assessment and evaluation. To roughly state, teachers perceived themselves knowledgeable regarding assessment purposes such as checking understanding as a result of teaching, identifying students' current knowledge before teaching, making decisions on students' success, informing students and other stakeholders about students' competency levels. Table 33 details EFL instructors' strengths related to purposes in assessment.

Table 33.

Teachers' Strengths in Assessment Purposes

Code	f	Code Description
Assessment for checking students' understanding	12	Perceived strength in assessment purposes. Assessing the students in order to check whether they have learnt the subjects or not.
Assessment for identifying students' current knowledge	2	Perceived strength in assessment purpose. Assessing students for identifying their pre knowledge.
Assessment for identifying attainment of objectives	2	Assess the students to see whether they have attained to objectives which were set at the beginning of the process, or not.
Assessment for deciding on students' success	2	Perceived strength in assessment purpose. Assessing students to have information which can be later used for making decisions on students' success.
Assessment for informing students about their levels	1	Perceived strengths in assessment purpose. Assessing students for identifying their competency levels.
Assessment for making students take lesson serious	1	Also, participant reported to assess his / her students for turning his/her lesson into a serious lesson.

Participants mentioned their various purposes in their assessment practices. Based on the analysis, their primary objective for assessing their students was found as checking students' level of understanding after teaching. Teachers' emerging perceived strengths are in line with their perception of assessment purposes therefore we can say that teachers' beliefs corroborate with their practices. The following excerpt clarifies teachers' strengths in assessing students to check their understanding.

*I mean, when you came to observe me, I was giving lectures at that time. If I had been to assess students at that moment, I would have assessed them to **check whether they understood what I was teaching, or not**. The questions I asked would have been prepared to determine whether the topics I taught in the class were understood or not. (Post semi-structured interview, P2)*

In addition, one of the purposes mentioned by the participants was to decide whether the students will pass the class or not. Although teachers were not observed in their classes while they are making decisions on the students' success, they strengthened their claims with their explanations in follow-up interviews. Based on the findings, we can say that teachers are competent in assessing their students and in turn making decisions on students' success on the basis of obtained results. The following excerpt exemplifies the situation:

*As for teacher's responses about **purposes of assessment**, s/he reported to **assess his/her students just for deciding on their success**. The participant explained that as a result of his/her assessment practices, s/he **makes decisions on whether students should pass the class or fail**. (Researcher reflection report).*

4.3.1.5. Teachers' Strengths in Ethical Considerations in Assessment

Participants in this study also answered questions about ethical issues during the interviews. As a result of the analysis, two sub-themes emerged. While one of the sub-themes provides teachers' strengths in perception and knowledge about ethical issues in assessment, the other one shows teachers' strengths in incorporating ethical issues in their assessment practices. Both of the sub-themes have been explained in the following parts.

4.3.1.5.1. Teachers' Strengths Regarding Views of Ethics

As mentioned earlier, this sub-theme is related to what participants think of when ethical issues are mentioned in assessment. Namely, it provides information about the participants' perceptions of ethics in assessment as it is detailed in Table 34.

Table 34.

Teachers' Strengths Regarding Views of Ethics

Code	f	Code Description
Associating ethics with objectivity	8	Perceived strengths in ethics – perception of ethics, objectivity.
Associating not including feelings into assessment	7	Perceived strengths in ethics – ideas on keeping personal relationships and feelings and thoughts in grading.
Associating fairness with ethics	5	Perceived strengths in ideas on creating fairness in assessment practices.
Associating ethics with equality	4	Perceived strengths in treating every student equally in assessment practices.

First of all, it should be noted that when the participants talked about ethics in assessment, they thought of administering exam and grading the students' performances. That is, they mostly mentioned the objectivity they provided while administering the exams and giving grades the students' answers and stated that they associated ethics with objectivity as it is shown in the excerpts:

*When I think of **ethical considerations in assessment, objectivity comes to mind.** (Pre semi-structured interviews, P1).*

*If there is **anything about ethical issues** for a while, I don't know about them. I just think that **I should be objective when giving grades.** (Pre-semi-structured interviews, P4).*

As can be seen from Table 34, participants' strengths in ethical issues in assessment differ in many ways. While some associated ethics with objectivity, others associated it with fairness, not including personal feelings in assessment practices, or equality. For example, being fair in assessment practices was found as a frequently mentioned expression by the teachers. The participants thought that teachers should be fair both in administering the exams and grading them. Furthermore, they claimed that they practiced as they believed in this regard. Although teachers were not observed while behaving fair and objective in class, through their exclamations in the follow-up interviews, they verified their perceived strengths regarding the issue. This result can be seen in the excerpts below:

*As I said, **when I say ethics, I think of fairness, that is, to be fair, to be ethical, for me, assessment and evaluation.** (Pre-semi-structured interviews, P3).*

*In terms of ethical considerations, the **participant associates ethics with being fair in terms of grading**. Apart from it, the participant reports that he/she didn't know extra things. (Researcher reflective report)*

4.3.1.5.2. Teachers' Strengths in Ethical Practices

It was also aimed to determine how the participants ensured ethics in their practices. Participants stated that they ensured ethics by considering factors such as not mixing their feelings and emotions in grading, being objective in grading, that is, making evaluations by treating each student equally, and being fair while evaluating students.

Table 35.

Teachers' Strengths in Ethical Practices

Code	f	Code Description
Being objective in grading for ensuring ethics	6	Strength in enhancing ethics in assessment practices through being objective in grading.
Ethical behaviour focusing on just objectives in grading	5	Strength in enhancing (implementing) ethics in assessment practices through focusing on goals and whether they are being achieved during grading.
Behaving equally to all students for ensuring ethics	4	Strength in enhancing ethics in assessment practices through behaving equally to all students both in teaching and assessment practices. Ethics in grading.
Using rubric while giving grades to ensure ethics	1	Strength in enhancing ethics in assessment practices through using rubric in grading. Therefore, participant believes that s/he will be ethical in his/her assessment practices.

Based on the explanations and the statements of the participants, it can be stated that participants' knowledge considering ethics was parallel to their actual classroom practices. In other words, the participants stated that they ensured ethics in assessment practices by behaving objectively while grading. Although this was not observed in the classroom observations, as a result of the analysis of the explanations made by the participants in the pre and post-interviews, it was inferred that the participants ensured the ethics in their assessment practices by acting objectively. The excerpt below exemplifies the situation:

*Well, for example, if the **teacher is not objective**, if the teacher **gives low grades to those who deserve high grades** and gives higher grades to those with whom s/he gets along well, this **behaviour is not ethical**. I always behave objectively while grading. (Pre-semi structured interview, P2).*

Additionally, the participants associated ethics with equality and they stated that they ensured ethics in their own practices by treating each student equally while giving grades to them. Participant 3 stated how s/he ensured ethics in his/her classroom practices:

*Now, when I think of ethics in assessment, I think of being equal while grading. We should treat every student **equally**. We should not give a student a high grade and a low grade for another student. **The same criteria** should be applied to all students. **Equality** comes to mind when I am said ethics. This is how I act when I give grades. (Pre-semi-structured interviews, P3).*

All these results show that the participating teachers' knowledge, beliefs and practices related to ethical issues in assessment are in line with each other.

4.3.1.6. Teachers' Strengths in Scoring and Grading

Another theme that emerged as a result of the analysis shows the perceived strengths of the teachers regarding scoring and grading in assessment. Two main sub-themes related to this theme emerged as a result of the analysis. Each sub-theme has been explained in detail in the next sections.

4.3.1.6.1. Teachers' Strengths in Grading

The first sub-theme is teachers' strengths in grading. Under this sub-theme, the self-perceived strengths of the instructors regarding grading, both in terms of knowledge and in-class practice, are brought together. To illustrate, teachers' main perceived strengths regarding grading were found as grading for finding out students' learning level, answer key usage, grading for decisions on students' success, being objective in grading. Table 36 presents the detailed content of this sub-theme.

Table 36.

Teachers' Strengths in Grading

Code	f	Code Description
Grading for presenting students' competency level	7	Strengths in purposes of grading. Participants express that they give grade in order to show students' successes or strengths along with their weaknesses and failures.
Using answer key in grading	5	Strength in practices considering grading. In order to better grade without measurement error, participants employ answer keys.
Grading for deciding on student's success	4	Strengths in purposes of grading. Participants express that they give grade in order to make decisions considering students' success. Based on grades, teachers decide on whether any student should pass the class or fail.
Ignoring quality of relationships with students in grading	3	Strength in practices considering grading. In order to give grades that reflect the real level, neither more nor less, the participants ignore their personal relationship with the students during the grading.
Being objective in grading for reliability	2	Strength in practices considering grading. In order to ensure reliability, in order not to include errors in grading, participants report to be objective in the process.
Using blind grading for reliability	2	Strength in practices considering grading. In order to ensure reliability, in order not to give extra grades, participants report not to look at the names of owners of the exam papers. That's why, they believe that they enhance equality and reliability.

First of all, most of the teachers touched upon their purposes for grading. The obtained results, which were found as in line with the previous findings of the study, appeared in two separate codes, one of which is grading for deciding on students' success and the other one is grading for demonstrating student' gains or mistakes. As their names suggest, these codes refer to why the participants actually give grades. Among the grading purposes, "Grading to decide on the success of the students" was found to be the most emphasized code in this section. With this code, the participants stated that they graded the students' answers in the exams or their performances numerically, and that these grades were then used as a criterion in determining making decisions about students' success. The excerpts below shed more light on the situation:

I give grades to determine the student's level numerically. I give grades to determine whether or to what degree the student has learned. Then, based on these grades, we decide whether students should pass or fail. We have to give grades for these decisions. I give grade to make decisions about student success. (Pre-semi-structured interview, P2)

She reported that she just gives grades in order to make decisions about students' competency levels. (Researcher reflection report)

In addition to the teacher's grading for making decisions on students' success, identifying students' competency levels was stated as another reason for their grading practices.

We assess and give grade to students' performances to find out whether they have the knowledge level necessary to pass the preparatory class, or not. (Pre-semi-structured interview, P3)

As this quote indicates, for the teachers, embodying students' competency levels was mentioned as their purpose for giving grades. Moreover, they stated that one of their strengths in grading was ignoring their personal relationships with the students and the degree of their sincerity with them during grading. As can be seen from this emerging finding, the participants cared acting objectively rather than being subjective in their grading processes. The excerpt below better illustrates this situation:

It is not possible to give a low or high grade to a student just because I have a good or bad relationship with the student. (Pre semi-structured interview, P3)

Last but not least, the participants indicated the ways to ensure reliability while grading under this sub-theme with two different codes. First of all, teachers paid attention to being objective, as they have mentioned before in ethical considerations. Participants think that the more objective they are in the grading, the more reliability will increase. Participant 2 stated how s/he enhances reliability in his/her assessment practices:

*My grading strength is **objectivity**. I ensure **objectivity** in my assessment practices. I think that I can evaluate exam papers effectively and accurately. **My strength** is to be **objective**. (Pre-semi-structured interview, P2)*

In addition to objectivity, using the blind grading technique while grading was also seen as a strength and applied by the participants as a way to increase reliability. To explain more, the teachers stated that while grading the answers on the exam papers, they only focused on the student's performance or the answers given by the student, without looking at the student's name, and naturally they thought that it increases the reliability. The excerpt below better exemplifies the situation:

*Since I started teaching, I never look at the student's name **while grading exams**. I do not consider the student's name. In this way, I ensure reliability in grading. (Pre-semi-structured interview, P2)*

4.3.1.6.2. Teachers' Strengths in Scoring

Throughout the study, perceived strengths of teachers concerning scoring in assessment were tried to be explored through both interviews and classroom observations. As a result, emerging strengths were found as scoring according to difficulty level of the exam, scoring by considering students' levels, and scoring according to focused parts during teaching. These strengths are shown in detail in Table 37.

Table 37.

Teachers' Strengths in Scoring

Code	f	Code Description
Scoring by giving high score based on difficulty level of item	2	Self-perceived strengths in considerations during scoring by focusing on the difficulty level of the test item. Assigning higher scores to the self-perceived difficult questions.
Considering students' level in assigning scores	1	Self-perceived strengths in considerations during scoring by focusing on students' level. Trying not to ask students too challenging and difficult questions or too easy questions for their competency levels. Getting to know about students' knowledge levels, then assigning scores.
Scoring considering parts which are more focused during teaching	1	Self-perceived strength in scoring by taking the topics taken care of to be taught seriously throughout the lessons.

First of all, it was revealed that teachers made arrangements in scoring of test items according to their difficulty levels. In other words, teachers assigned high scores to difficult questions and low scores to easy ones that every student could answer correctly. The excerpt below explains the situation:

*While scoring, I score according to the difficulty and ease of the test items. For example, I assign a high score to the questions that require the student to use more information namely to **difficult questions**. But I assign low scores to easy question types such as True / False. (Post-semi-structured interview, P2).*

In short, teachers perceive assigning higher grades to the difficult questions in the exams as their strengths in scoring.

Secondly, the participants stated that they assigned higher scores to questions including subjects that they focused more to help student learn them. In other words, scoring weight was

determined considering the subjects that the teachers gave importance to teach in their lessons. The following excerpt elaborates this situation:

When scoring, I consider the topics we cover in the lesson. For example, I assign higher points to the topics that I focus on during the lessons, the topics that we consider as important, that is, the questions that assess whether these topics have been learned or not. (Pre semi –structured interview, P3)

4.3.1.7. Teachers' Other Strengths Regarding Assessment

This emerging theme provides information about teachers' perceived strengths in assessment which were not included in above-mentioned themes. These strengths are classified into two sub-themes as teachers' other assessment related strengths and teachers' strengths in avoiding from measurement error. Both of them are explained in detail in the following sections.

4.3.1.7.1. Teachers' Other Assessment Related Strengths

The opinions of the participants belonging to this emerging sub-theme provide teachers' perceived strengths related to the reliability and validity in assessment, effects of the assessment practices on students, and the effects of valid evaluation. Table 38 portrays the sub-theme:

Table 38.

Teachers' Other Assessment Related Strengths

Code	f	Code Description
Considering content validity for validity as strength	11	Strength considering validity in general by focusing on content validity.
Influence of assessment practices on students.	3	Strength in explaining the influence of assessment practices on students
Considering objectivity for reliability	3	Strength considering reliability in general by only focusing on being objective.
How valid evaluation influences strength	1	Strength in explaining the possible influences of valid evaluation on decisions related to instruction.
Strength summative assessment type	1	Strengths in summative assessment type.

When the answers given to the interview questions were analysed, the only strength of the participants regarding validity in assessment was content validity. This code was the one with the highest frequency. Each time the participants talked about validity, they only

mentioned content validity. Therefore, participants' strength regarding validity can be about only content validity. The excerpt below illustrates the situation better:

*Validity is the assessment of the subjects that we teach students in exams. **Not including extra topics other than these topics in the exams.** That is, validity is the questioning of what has been taught. I only know that when it comes to validity. (Pre-semi-structured interview, P3).*

The strengths of the participants related to reliability were also found in the same direction as it was in validity. That is, the participants only talked about objectivity while answering questions about reliability in the interviews. In other words, the participants' knowledge of reliability as a concept, namely their strengths, was limited only to objectivity. Participant 2 stated his /her strengths in reliability by touching upon objectivity:

*Validity and reliability... **By reliability** I mean how **objective** it is, how **impartial** it is. (Pre-semi-structured interview, P2).*

Some of the questions asked to the participants were about the influence of assessment practices on students. Participants stated that assessment practices had only an effective influence on students. In other words, students took the course seriously thanks to the assessment and evaluation practices, and these thoughts were found as their own perceived strengths related to this subject. The excerpt from the interview below explains this situation:

*In other words, I think it has a positive effect. Since **there will be an exam**, the student **takes the lesson seriously** or pays more attention to the lesson. In other words, the student gives importance to the exam and I realize this. (Pre-semi-structured interview, P1).*

4.3.1.7.2. Teachers' Strengths in Avoiding from Measurement Error

Finally, in this sub-theme, teachers' perceived strengths in avoiding measurement errors in their assessment applications are presented. Participants considered the practice of avoiding measurement errors when they were only administering and grading exams. Table 39 provides explanatory information about these self-perceived strengths of the teachers.

Table 39.

Teachers' Strengths in Avoiding from Measurement Error

Code	f	Code Description
Ensuring objectivity in preventing measurement error	2	Strengths considering the ways of avoiding from measurement error. Here participants indicate that enhancing objectivity in grading can be a good way.
Giving clear instructions to students to prevent measurement error	1	Strength considering interfering measurement error in assessment practices. Here participant indicates that by giving clear instructions, errors cannot be included in the assessment procedure.
Preventing cheating to avoid from measurement error	1	Strength considering interfering measurement error in assessment practices. Here participant indicates that by hindering cheating during the exam, measurement error can be avoided from.
Using answer key for preventing measurement error	1	Strength considering preventing inclusion of measurement errors into assessment procedure. It is believed that measurement error can be prevented through use of answer key during grading.

As can be seen from the Table 39 above, teachers reported to prevent the inclusion of measurement errors in their assessment practices by behaving objectively while giving grades to students' answers or performances. While grading, teachers think that by treating everyone equally and providing objectivity, they will prevent the inclusion of errors in the assessment. The following excerpt explicates the notion:

*In other words, since I mostly associate the measurement error with the objective behaviour of the teachers while grading the exams, as I said, **I try to be objective while grading**. I think that no mistakes are involved in my assessment practices. (Pre-semi – structured interview, P1).*

Again, as it is mentioned before, the participant associated his/her strength in this topic with the precautions s/he takes while applying the exams. To put it simply, the participant stated that s/he never allowed cheating in his/her exams in order to prevent the inclusion of errors in his/her assessment practices. Participant 3 explained how s/he ensured this in his/her practices below:

*Since I associate **measuring error with cheating**, I prevent **cheating** in exams. In this way, I think that there are no errors in my applications. (Pre semi-structured interview, P3).*

In addition, the participants stated that they prevented inclusion of the measurement error in their applications by giving clear instructions to students during the exam, using an answer key in grading and they considered themselves as strong in these areas. The statements

of the participants regarding these self-perceived strengths are explained in detail in Table 39 above.

4.3.2. Teachers' Weaknesses Related to the Assessment

As stated before, through this investigation, it is aimed to reveal the participating EFL instructors' weaknesses considering language assessment. The strengths of the participants were explained in detail in previous sections. Their weaknesses are explained in the rest of this part. First of all, in order to determine the weaknesses of the participants considering assessment especially language assessment, a pre-semi-structured interview, a classroom observation and then a post-semi-structured interview were applied to each participant. Meanwhile, the researcher's reflection reports were included in data collection tools. First of all, when determining the weaknesses of the participants, both the subjects about which participants directly declared themselves as weak and the incomplete and incorrect answers that they gave to the questions were taken into consideration. Then the obtained data were subjected to content analysis using Maxqda20 software. As a result of the analysis, eight broad themes related to teachers' weaknesses in assessment emerged. These findings are shown in Table 40.

Table 40.

Teachers' Weaknesses Related to the Assessment

Themes	Subthemes
Teachers' weaknesses related to test construction and test items	Teachers' weaknesses in test construction Teachers' weaknesses in item analysis Teachers' weaknesses in including test items in a test Teachers' weaknesses in item writing Teachers' weaknesses in the purpose of item usage
Teachers' weaknesses about alternative assessment	Teachers' weaknesses in knowledge of alternative techniques Teachers' weaknesses in performance assessment and rubric Teachers' weaknesses in practicing alternative assessment techniques Teachers' weaknesses in alternative assessment
Teachers' weaknesses in introduction topics in assessment	Teachers' weaknesses in assessment-related definitions Teachers' weaknesses in assessment types Teachers' weaknesses in purposes of assessment
Teachers' weaknesses about assessment	Teachers' weaknesses in interpreting assessment results

results	Teachers' other weaknesses related to assessment results Teachers' weakness in using assessment results
Teachers' weaknesses related to the assessment technique	Teachers' weaknesses in knowledge and implementation of assessment technique Teachers' weaknesses in criteria for technique selection Teachers' weaknesses in the purpose of technique usage
Teachers' weaknesses in grading and scoring	Teachers' weaknesses in grading Teachers' weaknesses in scoring
Teachers' additional weaknesses in assessment	Teachers' weaknesses about ethics in assessment Teachers' other assessment related weaknesses
Teachers' weaknesses regarding assessment principles	Teachers' weaknesses in validity Teachers' weaknesses in reliability Teachers' weaknesses in influence of assessment

When the Table 40 is investigated, it can be seen that teachers' weaknesses regarding language assessment have been classified under seven themes such as teachers' weaknesses related to test construction and test items, teachers' weaknesses about alternative assessment, teachers' weaknesses in the introduction topics in assessment, teachers' weaknesses about assessment results, teachers' weaknesses related to the assessment technique, teachers' weaknesses in grading and scoring, teachers' additional weaknesses in assessment, and teachers' weaknesses regarding assessment principles. Each theme has been explicated in the following sections.

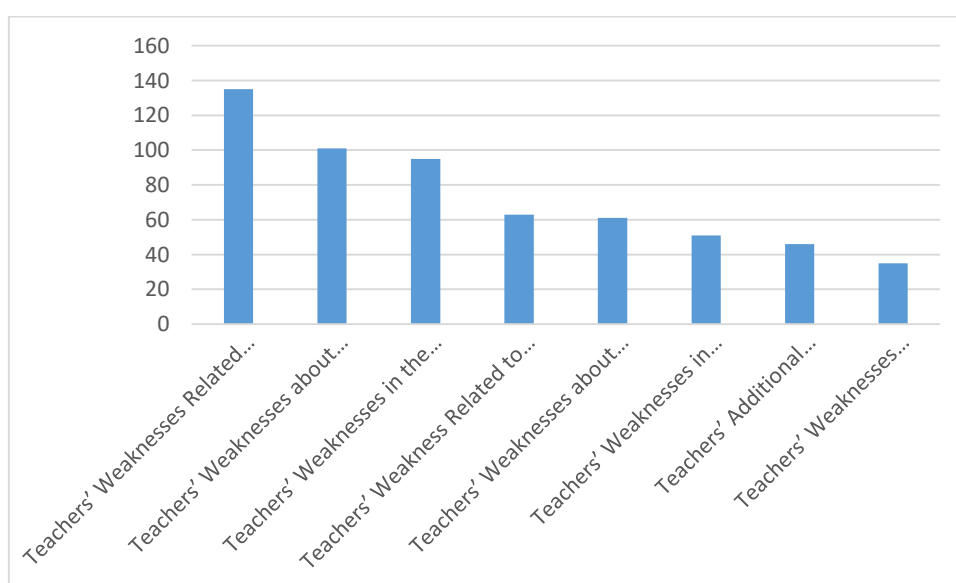


Figure 5. *Teachers' weaknesses Related to the Assessment*

Before the RRD sessions, teachers were subjected to various data collection tools to identify their weaknesses, namely their lack of knowledge about language assessment. In the light of the data obtained, the content of the RRD sessions was determined to a large extent. Teachers' weaknesses about language assessment were grouped under eight themes. The main subject areas in which teachers lack knowledge were test construction and writing test items, alternative assessment, assessment results, assessment techniques, grading and scoring, assessment principles. While all of the themes that emerged as a result of the analysis showed the weaknesses of the teachers in language assessment, the theme with the highest frequency of citation among them was found to be the area referring to the findings related to the teachers' lack of knowledge about test construction and writing test items. All the weaknesses of teachers related to assessment were detailed in the following sections.

4.3.2.1. Teachers' Weaknesses Related to Test Construction and Test Items

As a result of the content analysis, it was found that the participants' primary perceived weaknesses were related to the test preparation and test items. With this theme, teachers' perceived weaknesses regarding the titles such as test construction and test items are presented. As a result of the analysis of the answers given by the participants to the interview questions, five broad sub-themes were determined. These are teachers' weakness in test construction, teachers' weakness in item analysis, teachers' weaknesses in including test items in a test, teachers' weaknesses in item writing, and teachers' weaknesses in the purpose of item usage. In the following sections, these sub-themes belonging to this emerging theme are discussed in detail.

4.3.2.1.1. Teachers' Weaknesses in Test Construction

First of all, teachers' weaknesses in test constructions were gathered under this sub-theme. To illustrate, this emerging sub-theme provides findings referring to teachers' weaknesses in knowledge of test construction and types of test formats, ensuring validity and reliability while constructing tests, and some practices performed during tests construction such as deliberately asking too hard and easy questions and helping students get higher grades through tests. These codes are detailed as they can be seen in Table 41.

Table 41.

Teachers' Weaknesses in Test Construction

Code	f	Code Description
Inadequate knowledge considering test preparation	15	Perceiving themselves as incapable of preparing tests. Perceived weakness in test preparation in general.
Ensuring validity in test preparation	12	Believing to have limited knowledge in validity especially ensuring validity in test preparation. Perceived weakness considering enhancing validity in test preparation.
Enhancing reliability in test preparation	9	Self-perceived weakness in reliability in test preparation. Perceiving their knowledge considering reliability in test preparation as insufficient.
Weakness in test formats	5	Perceived weakness in test formats like subjective vs objective test, their characteristics, types of test items to be included in those tests.
Asking easy questions in test	4	Self-perceived weakness in preparing test considering the item difficulty level of the test items. Being a supporter of helping students for getting higher grades. Choosing or writing easy questions and include them in test.
Including various kinds of items in test prep	3	Perceiving themselves as having limited knowledge in integrating various kinds of test items in a single test.
Regarding students' getting high grades while preparing tests	2	Exclaiming that they act consciously in order to help students get higher grades and prepare their test accordingly.
Exam preparation process in practice	2	Explaining the weaknesses in the process of test preparation.
Asking hard questions in test	1	Expressing consciously to write or select hard questions for the test.

First of all, the participants directly stated that they did not know how to prepare a test in general although they were familiar with the terms included the interview questions. They reported to either prepare it by heart or download a sample of it directly from the internet without knowing nothing about what they should do while preparing the test. The following excerpt illustrates this:

I use textbooks for mid-terms. I prepare an exam by copy-pasting ready-made questions from the teacher's book. However, there are extra activities I give to students, I include questions similar to those activities in the exam. (Pre-semi-structured interview, P3).

Then, the participants were asked questions about how they ensured validity and reliability while preparing the test. The answers given by the teachers actually were found to be in line with the findings showing participants' strengths in previous parts of the current study. In other words, comparing the two sets of results, it can be seen that participants had

limited knowledge considering validity. As they stated before, content validity came their minds when they were asked to tell something about validity. Overall, these results indicate that participants had weaknesses about ensuring both validity and reliability in test construction. The following excerpts illustrate the participants' weaknesses in terms of ensuring validity and reliability in test construction:

I'm sure, I have a weakness, a lack of knowledge, to ensure validity in test preparation. My assessment and evaluation knowledge depends entirely on the practices I have been exposed to so far. Surely there are different ways, but I do not know them. As I said before, I know content validity. (Pre-semi-structured interview, P3).

When I talk about my weaknesses, that is, if someone asks me, "What do you do to ensure reliability in the exam?" I cannot answer it. I prevent cheating while applying the exam, but I don't know how to ensure reliability while preparing a test. (Pre-semi-structured interview, P1)

Interestingly, two of the participants mentioned that while preparing the exams, they aimed at making them easy for students so as to help them get higher grades. For this reason, they honestly explicated that they deliberately included easy questions in the tests. Based on these explanations, it was seen that the weak points of the participants in this regard were that they allowed students to get high grades while preparing the test, rather than assessing the students, and that they deliberately included easy questions in the test for this purpose. Based on these explanations, one of the participants' statements illustrate the situation:

I take care to prepare an exam that will not be too difficult for the students and from which almost everyone can get high grades in some way. (Pre-semi-structured interview, P1).

In addition, the teachers stated that they did not know the types of tests such as subjective tests or objective tests and they included tests similar to the ones they had been exposed to and used for years while preparing the exam. It was found that teachers prepared exams using question types derived either from their familiarity during their student years or from their teaching experiences, without knowing which question type they included and why. On the basis of these results, we can infer that teachers have weaknesses related to test preparation.

4.3.2.1.2. Teachers' Weaknesses in Item Analysis

This current sub-theme provides results referring to teachers' perceived weaknesses concerning item analysis. These weaknesses were found as teachers' weaknesses regarding knowledge of item analysis, conducting items analysis based on obtained assessment results, and item analysis in test construction. Table 42 details teachers' weaknesses in item analysis.

Table 42.

Teachers' Weaknesses in Item Analysis

Code	f	Code Description
Knowledge of item analysis	17	Weakness regarding knowledge of item analysis. Although participants report to remember this subject from their university courses, now they do not know it enough.
Item analysis based on assessment results	15	Participants reported that they interpret the assessment results without analysing the items such as their items discrimination level, item difficulty level. They do not know how to do this.
Item analysis in test construction	4	Weakness in implementation of item analysis in test construction. During test construction process, participants reported not to analyse the test items and then put them into the test.

When Table 42 is examined, teachers' weaknesses related to item analysis are classified under three different parts, such as lack of knowledge in item analysis, use of item analysis in assessment results, and implementation of item analysis in test construction. Even though the participants partially remembered the item analysis from their university courses, they admitted their weakness and stated that the reason for this was that they hadn't been required to use item analysis in their practice so far and therefore they did not need to learn it. As for the integration of item analysis into assessment results, the participants declared that they did not know any statistical calculations to analyse the assessment results, and they made hypothetical interpretations by looking at the results. Finally, participants came up with statements that were in line with the answers they gave in previous sections to the questions about test construction. In other words, they reported that they did not have enough information about the test preparation process and therefore they did not know how to include item analysis in this process, they constructed their tests by heart, without considering anything, or they just downloaded them from the internet. Based on these statements, it can be inferred that the participants have weaknesses that need to be developed regarding item analysis.

4.3.2.1.3. Teachers' Weaknesses in Including Test Items in a Test

This sub-theme shows teachers' weaknesses regarding selection and inclusion of various types of test items for a single test which is appropriate for their assessment

objectives. To illustrate, teachers were found as having weaknesses in what to consider while choosing test items for a test, in knowledge of different kinds of test items. They also were found to include test items with which students were familiar. Teachers' weaknesses concerning this issue are portrayed in Table 43:

Table 43.

Teachers' Weaknesses in Including Test Items in a Test

Code	f	Code Description
What to regard in selection of test items for a test	7	Weakness in considerations in item selection for a test. Here, participants report that they do not know what take into considerations for choosing or writing suitable test items in a test.
Weakness in including suitable items in relevant test formats	5	Participants reported that they do not know types of test formats. Based on this, they also report that they do not know what to do while choosing suitable test items according to test formats. Self-perceived weakness in item selection for test formats.
Kind of test item to be included in a test	4	Participants express that they do not well about types of items, how and for what purposes they should be included in a test. Self-perceived weakness in integration of various kinds of test items and considerations for their inclusion.
Including test items in a test by heart	4	Weakness in inclusion of test items in a single test considering assessment objectives. The participant stated that she did not consider anything while including a test item in any test, and included the items in the test according to her habits or randomly.
Including familiar questions in a test	1	Weakness in integration test item types in a test. Participant reports that s/he asks consciously questions with which students have been acquainted with in order to help them in exams.

First of all, it can be deduced that two weaknesses which came out as a result of the analysis are closely related to each other. To explain this situation in a detailed way, based on the explanations of the participants, we can say that teachers did not know what to consider when writing a test item or when choosing a test item for a test. That's why, the participants stated that they acted by heart while implementing these things. The excerpts exemplify the situation:

*I usually use question types such as fill in the blanks, matching, open ended. **The reason I use them is to provide diversity.** I've seen them in exams for years. That's why, I'm asking or using them. Other than that, I don't know what and why I should ask. (Pre-semi-structured interview, P3).*

*I don't know why I use various question types. **These are works done by heart.** Also, adding only one question type to a written English test is not enough. I add more than one type of question for diversity. (Pre-semi-structured interview, P2)*

Additionally, we can also say that teachers do not have sufficient knowledge about test item types. In other words, the teachers preferred directly matching, true/false, fill in the blanks and multiple-choice question types while writing test items or choosing them from the internet. It was also shown that teachers did not know how to apply Bloom's taxonomy when deciding on the test item, and they used these question types just for diversity. From all these explanations, the emerging results suggest that the teachers did not have sufficient knowledge about the types of test items. Although they incorporated them in their test construction processes, they did it without knowing appropriate expertise. The excerpts provide better explanations about the issue.

*I choose or write a test item to **determine whether students have learned and to grade students**. Then I construct a single test from them. I don't know what type or format of this test is. Maybe I do, I'm not sure. (Pre-semi-structured interview, P4).*

*As I mentioned, **I do not know the types of test items**. These are my shortcomings and I noticed this shortcoming when you asked. (Pre-semi-structured interview, P1).*

Similar to the interesting finding appeared in the previous sub-theme, one of the participants stated that s/he consciously included questions or test items that students were familiar with in the test to help them demonstrate better performances during the exam. Considering these, we can conclude that the participant teachers need development in including relevant test items in an appropriate test.

4.3.2.1.4. Teachers' Weaknesses in Item Writing

During the interviews, teachers were also provided with questions about writing test items. Three different codes emerged from the answers such as what to consider in item writing, considering Bloom's taxonomy in item writing and teachers' deficiency in writing multiple test items and they constituted the emerging sub-theme. You can find these codes, their frequency and descriptions from Table 44:

Table 44.

Teachers' Weaknesses in Item Writing

Code	f	Code Description
Insufficient knowledge of what to regard in item writing	6	Weakness in item writing. Participants report that they do not have enough knowledge of considerations for item writing.
Bloom taxonomy in item writing	5	Self-perceived weakness in Bloom's taxonomy. Participants report that they do not well Bloom's taxonomy and that they cannot take it into consideration while writing test items.
Inadequate knowledge in writing multiple choice questions	4	Participants perceive that they do not know the principles for writing multiple-choice items. They copy from the internet and paste them into their exams or they write them directly without taking care of any rules.

It was deduced from the answers given to the interview questions that participants did not have sufficient knowledge of writing test items. For example, when asked about Bloom's taxonomy, the participants stated that they neither knew about Bloom's taxonomy nor they took this into account while writing the tests items. Furthermore, participants confessed that thanks to these questions, they first learned that Bloom's taxonomy could be used when writing test items. The following extracts exemplify participants' explanations:

*Yes, I remember a little bit about Bloom's taxonomy, but **I can't remember the steps of the taxonomy.** I haven't not included it in the questions I prepared. **I can't. I don't know how to include it.** (Pre-semi-structured interview, P1)*

*As I write test items, **their content should be in line with the topics I teach.** I don't ask questions whose subjects haven't been taught by me. I also use understandable expressions while preparing the questions. My instructions are clear. Apart from these, what can I consider when I am writing a test item? -I don't know. (Pre-semi-structured interview, P3)*

It is understood from the quotes above that the participants needed development on the subject of the writing test item.

Moreover, questions related to writing multiple choice question type, which is a specific test item, were asked to the participants. From the answers of the participants, they were found to mostly use multiple choice test item type and they did not know how to write it. What they did was only considering the content of the test item and copy it from the internet or other sources. The fact that teachers only paid attention to the content while creating multiple choice test items was in line with their validity-related weaknesses as it was reported

that they did not have detailed knowledge in validity except from content validity. The following excerpt exemplifies teachers' weakness in writing multiple choice test.

*I heard that it is very difficult to prepare multiple-choice exam questions. I don't prepare it either; **I download it from the internet. I pay attention to whether the question is too difficult or too easy.** If it is also relevant to the subject I teach, I will copy and paste it. (Pre-semi-structured interview, P2)*

4.3.2.1.5. Teachers' Weaknesses in the Purpose of Item Usage

For this sub-theme, the participants made explanations about their purposes of using test items that they included in their tests. These explanations provide information about what kind of test items teachers generally used and for what purposes they employed these items. The participant teachers' answers showed that they included the test items they claimed to use in their assessment practices by having inappropriate purposes. For instance, Table 45 details this sub-theme:

Table 45.

Teachers' Weaknesses in the Purpose of Item Usage

Code	f	Code Description
Including test items to help students get high grades	5	Participants explained their purposes for including any kind of test item in a single test. Perceived weakness in the usage of test items in a test.
Weakness in explanation of including various test items in a test	3	Weaknesses in explaining the reason for using any kind of test item.
Inadequate knowledge of purpose of including T/F test item	1	Weakness in writing and using T/F test items.
Weakness in purpose for using a multiple-choice test item	1	Weakness in writing and using multiple-choice test item.

While talking about the purpose of including any test item in a test, it was determined that teachers included test items in their exams to ensure that students get high grades, rather than revealing the real competency levels of the students. The following excerpts clarify the participants' aforementioned purpose:

*I add a variety of test items that allow **students to take the exam more easily and get higher grades.** For this reason, I often include the True / False question type in my exams. (Pre-semi-structured interview, P1)*

From the purposes and quotations above, it can be understood that the participants had inappropriate intentions in choosing the test item and including it in the test.

Next, some of the weaknesses frequently mentioned by the participants were also related to the inclusion of various types of test items in the exams. Although the participants stated that questions in the exams should consist of different types of test items rather than focusing on a single type, they could not explain the reason for doing this and accepted it as their self-perceived weakness. In addition, the participants explained that they implement including various kinds of test items by heart to create diversity and difference. The participant explains how s/he thinks about enhancing diversity in terms of test item types in the exams:

Frankly, I always add matching, fill-in-the-blanks, multiple choice and open-ended question types to exams. I add true and false because the student can guess the answer and get high grades. There are questions that I ask so that the students can get higher grades so that the exam will be easy. For example, True / False questions. I definitely provide diversity in exams, but I guess I can't fully explain the reason for providing this diversity. As I said, the exam should be seemed good; the students should not get bored while answering the questions in the exam. (Pre-semi-structured interview, P3)

4.3.2.2. Teacher's Weaknesses about Alternative Assessment

Teachers' answers to the interview questions revealed that they did not have enough knowledge about alternative assessment. Participating instructors were generally found to prefer traditional assessment in their practices. Similarly, teachers indicated that they had weaknesses concerning both knowledge and implementation of various alternative techniques and the advantages offered by alternative assessment. Four sub-themes such as teachers' weaknesses in knowledge of alternative techniques, teachers' weaknesses in performance assessment and rubric, teachers' weaknesses in practicing alternative assessment techniques, and teachers' weaknesses in alternative assessment emerged as a result of the analysis. They have been explained separately in the following sections.

4.3.2.2.1. Teachers' Weaknesses in Knowledge of Alternative Techniques

In the interviews, participants were first asked what they generally know about alternative assessment. These questions were roughly like this: "Why should alternative assessment be used? or How did alternative assessment come out?". Then, general questions

about alternative assessment techniques were asked of the participants. Afterwards, the teachers were provided with questions focusing on specific alternative techniques. From the answers given by the participants to such questions, their weaknesses in the relevant subjects were determined. To go into detail, it was found out that teachers did not have enough knowledge about alternative assessment techniques such as portfolio, self-assessment, peer assessment, and classroom observation. As a result, when they tried to apply them, they applied them incorrectly or they did not apply them at all in their classrooms. Table 46 gives detailed information about this sub-theme:

Table 46.

Teachers' Weakness in Knowledge of Alternative Techniques

Code	f	Code Description
Insufficient knowledge of portfolio assessment	15	Weakness of the participants in the knowledge level of the portfolio assessment.
Insufficient knowledge of self-assessment	13	Weakness in the knowledge level of participants about self-assessment
Weakness in knowledge of kinds of alternative techniques	12	Self-perceived lack of knowledge the participants about alternative assessment techniques
Insufficient knowledge of peer assessment	3	Weakness in the knowledge level of participants about self-assessment
Insufficient knowledge of classroom observation	1	Weakness of the participants in the knowledge level of the portfolio assessment

Based on the results, the participants' weaknesses regarding knowledge about specific alternative techniques were gathered around portfolio assessment, self-assessment, peer assessment and classroom observation. Participants mentioned several times that their knowledge about portfolio assessment was limited to hearsay information from the environment and they did not know enough about the portfolio assessment process in order to include it in their practices. Based on this, it can be said that the participants did not know enough about portfolio assessment. The following excerpts from the interview illustrate the situation:

*I do not know **portfolio assessment** exactly, so I do not know much about **how to apply it to students**, and how to maintain its application throughout the process. (Pre-semi-structured interview, P1).*

*Because I don't know exactly **how to implement a portfolio**. I don't want to be embarrassed towards students either. That's why I intend for a portfolio assessment every year because **I always hear that it works very well**, it's process-oriented or*

*something. But since **I don't know, I can't apply it fully.** (Pre-semi-structured interview, P3).*

Secondly, the participants were asked whether they were familiar with the types of alternative assessment techniques in general, or not. Based on the answers, it can be put forward that the alternative technique knowledge of the participants was limited because all of the teachers answered the question just focusing on portfolio and self-assessment. They were also found to be incomplete. Apart from them, they did not talk about any techniques. The excerpts below illustrate the situation:

*There are many **contemporary assessment methods. I don't know them.** I am familiar with them by name, **but I don't have enough knowledge** of what they are, what their advantages are, for what purposes, in which courses, with what kind of students they are used, how they are applied. (Pre-semi-structured interview, P1).*

Last but not least, the participants also mentioned their weaknesses in knowledge about self-assessment, which is one of the alternative techniques. To explain this situation more specifically, the participant teachers stated that they were exposed to self-assessment when they were students at university, since then, they had not implemented it in their practices because they did not have enough knowledge to implement it in their own assessment practices.

*Now, to tell the truth, I can say its Turkish meaning, is self-evaluation. I was only exposed to such a practice when I was at the university in Eskişehir, but our assessment applications were always in the form of written exams like mid-terms and finals. Therefore, **I don't know what self-assessment is.** Now, if we leave the student alone with himself, how the objective will the student be in evaluating himself? The student gives himself higher grades, so I don't apply it coz I don't know. (Pre-semi-structured interview, P2).*

These emerging weaknesses were found to be in line with the teachers' findings in the previous research question. In other words, it was found that most of the teachers perceived assessment as making decisions about students' success by focusing on the result obtained at the end of teaching. In addition, it was found that the assessment techniques that most participants knew and included in their practice were written and oral exams. Also considering the weaknesses of the teachers regarding alternative assessment techniques, we can say that they adopt a traditional assessment before the RRD sessions.

4.3.2.2.2. Teachers' Weaknesses in Performance Assessment and Rubric

This sub-theme provides the weaknesses of the participating instructors regarding performance assessment, rubric design, and incorporation of performance assessment along with rubrics into participants' assessment practices. As a result of the analysis, it was found that the teachers knew the performance assessment incorrectly, they did not use rubrics in their applications, and when they needed to use it, they downloaded it from the internet and used it without adapting it. All these points out that teachers have weaknesses in performance assessment knowledge and application, rubric design, its adaptation and application. The following Table 47 presents elaborates this sub-theme.

Table 47.

Teachers' Weaknesses in Performance Assessment and Rubric

Code	f	Code Description
Weakness in the design of the rubric	13	Self-perceived weakness in rubric design. Participants reported not having enough information about
Weakness in knowledge of performance assessment	8	Participants reported that they do not have enough information about performance assessment. Self-perceived weakness in performance assessment.
Not employing rubrics in assessment practices.	8	The rubric is not preferred by the participants in assessment and evaluation practices
Not practicing performance assessment	5	Performance evaluation is not used much in actual classroom assessment and evaluation practices of the participants.

When explaining this sub-theme, it would be useful to mention rubric design and its usage separately, performance assessment and its implementation separately. Each aforementioned part actually contains two codes. Since both of these two codes were thought to be related to each other, we decided that it would be useful to consider them together and explain them accordingly.

First, the questions were about the knowledge of performance assessment and its use in practice. When the participants were told about performance assessment, they were aware that the student had to perform something. Their knowledge of this subject was limited to only this detail. In addition, it was understood from their statements that the participants associated the performance assessment with the speaking skill and the exams in which this skill was assessed. Except for those mentioned, the participating instructors did not add any detail about performance assessment. Therefore, it can be said that their knowledge about performance assessment was limited. The quotations directly taken from the interviews below exemplify the views of the participants:

When it comes to performance assessment, I can say that it is a type of assessment in which the performance of students is measured. Speaking skill comes to my mind. I don't know that actually. I've never used it until now. I don't know what it is, how it is planned, how it is implemented. (Pre-semi-structured interview, P2).

*In other words, when **performance assessment is mentioned, I think of assessing the student's speaking performance.** I set a few criteria in my head such as fluency, accuracy, grammar, vocabulary choice, pronunciation, and I divide them into segments. So that's how I implement it. (Pre-semi-structured interview, P3).*

For the questions asked about the use of performance assessment in the practices, three of the four participants clearly remarked that they did not use it. At the same time, since these participants reported not knowing performance assessment, it can be inferred that they did not use this assessment technique in their practice due to their lack of knowledge. The excerpts presented below show that participants do not utilize performance assessment in their practices:

*Well, I don't assess performance at school. We assess reading, grammar and vocabulary in exams. **I don't assess speaking. I usually assess reading. Because I have reading classes. So I don't use performance assessment.** (Pre-semi-structured interview, P4).*

Regarding the rubric, the participants also declared that they did not know how to prepare a rubric, and they didn't include it in their practice either. When they were asked to use it, they utilised the rubric by downloading it directly from websites they deemed reliable on the Internet, without doing any research.

*So how is the rubric prepared? **I don't know how to prepare a rubric for performance assessment.** That is, what is its content, what are its types, how is its usage? (Pre-semi structured interview, P1).*

4.3.2.2.3. Teachers' Weaknesses in Practicing Alternative Assessment Techniques

This sub-theme informs that participants had weaknesses in terms of implementing some specific alternative assessment techniques. In one of the aforementioned sub-themes belonging to the theme of perceived weakness about alternatives, it was revealed that the instructors had limited knowledge about alternative assessment techniques. Additionally, they stated that they did not know some specific techniques. Therefore, they naturally reported not being able to use them in their assessment practices. Therefore, this sub-theme focuses on the

weaknesses of the participants within the framework of the use of alternative assessment methods by centring on some specific alternative techniques. Table 48 details the sub-theme:

Table 48.

Teachers' Weaknesses in Practicing Alternative Assessment Techniques

Code	f	Code Description
Not practicing portfolio assessment	7	Self-perceived weakness in practicing portfolio assessment- not using it due to lack of sufficient knowledge about its implementation.
Not practicing self-assessment	4	Self-perceived weakness in practicing self-assessment- not using it due to lack of sufficient knowledge about its implementation.
Misuse of self-assessment	4	Self-perceived weakness in practicing self-assessment- applying it wrongly due to lack of sufficient knowledge about its implementation.
Not practicing classroom observation	1	Self-perceived weakness in practicing classroom observation- not using it due to lack of sufficient knowledge about its implementation.

This sub-theme focuses on weaknesses of the teachers concerning use of alternative assessment techniques focusing some specific ones. According to results, the participants explicated that they could not naturally use some specific alternative techniques since they were not knowledgeable enough about them. Participant 4's explanation exemplifies the notion: *Unfortunately, I have never used portfolio in the classroom so far. I have memorized information about it. If we know something, we can apply it. I do not know how to create the portfolio, how to manage the process, how to plan it. (Pre-semi-structured interview, P4)* Participant 1 also expressed in the following sentence that s/he does not include self-assessment, which is a special alternative assessment technique, in her practices: *No, I do not apply self-assessment to my students. (Pre-semi-structured interview, P1)*

4.3.2.2.4. Teachers' Weaknesses in Alternative Assessment

The least frequently mentioned sub-theme, weakness of alternatives in general, reveals the participants' weaknesses in knowledge about alternative assessment in general. These self-perceived weaknesses were found to be related to the emergence of alternative assessment, for what purposes it is utilized and the advantages that it is provided with. Table 49 presents more detailed information about this sub-theme:

Table 49.

Teachers' Weaknesses in Alternative Assessment

Code	f	Code Description
The perceived weakness of why to use alternatives	4	Participants stated why they should use alternative assessment methods, in other words, they have insufficient knowledge about the advantages of alternative assessment methods.
The perceived weakness of strengths of alternatives	2	Participants stated that they had insufficient knowledge about why they should use alternative assessment methods.
Misconception about alternatives- focus on the process	1	It is thought that the participants have a wrong perception of alternative assessment. Participants perceive alternative assessment as an assessment that focuses only on the process.

First of all, looking at Table 49, it became clear that teachers did not know why alternative assessment should be included in assessment practices. In fact, we can say that these findings were found to be in line with each other. For instance, as a result of the analysis, we found that the teachers also did not know the advantages and benefits offered by alternative assessment. In short, teacher knew neither why to implement alternative assessment nor advantages provided by it. Then, we can say that these two results can actually be associated with the fact that teachers did not include alternative assessment in their assessment practices. In short, these codes give the message that teachers had insufficient knowledge about alternative assessment, advantages that alternative assessment offers to teachers and students, and why it should be included in assessment practices. Following excerpt explicates the current finding:

*It was observed that teachers **did not apply alternative assessment in their practices**. In the interviews, it was seen that **the teachers did not know why alternative assessment should be included** in the practices, what it was for or the advantages it offered. For these reasons, it can be said that teachers do not apply alternative assessment. (Researcher reflection report)*

4.3.2.3. Teachers' Weaknesses in Introduction Topics in Assessment

In the interviews conducted, the participants were asked questions about the basic issues in assessment and evaluation. As a result of the analysis of the answers, their relevant weaknesses were determined. These weaknesses are categorized under three headings such as teachers' weaknesses in assessment-related definitions, teachers' weaknesses in assessment types, teachers' weaknesses in purposes of assessment. In the following sections, these sub-themes have been explained one by one.

4.3.2.3.1. Teachers' Weaknesses in Assessment-related Definitions

Questions related to some definitions of assessment were added to the interview questions. In addition, the participants sometimes added some definitions to their answers while they were talking. As a result of the analysis, the weaknesses of the participants related to these definitions were revealed. The main definition-based weaknesses are related to measurement error, assessment, measurement, evaluation, reliability, validity, consistency which are shown in Table 50.

Table 50. *Teachers' Weaknesses in Assessment-related Definitions*

Code	f	Code Description
Incomplete knowledge of measurement error	13	Self-perceived weakness about definition of measurement error
Weakness in definitions of criterion vs norm referenced. assessment	9	Self-perceived weakness regarding criterion referenced and norm referenced assessment definitions
Limited knowledge in definition of assessment	8	Self-perceived weakness regarding assessment definition
Weakness in definition of valid assessment result	7	Self-perceived weakness in definition of valid assessment result
Weakness in definitions of grading and scoring	6	Self-perceived weakness in definitions of scoring and grading.
Weakness in definition of reliability	6	Self-perceived weakness in definition of reliability
Weakness in definition of validity	5	Self-perceived weakness in definition of validity
incomplete definition of consistency	5	Self-perceived weakness in definition of consistency
Weakness in definition of measurement	3	Self-perceived weakness in definition of measurement
Weakness in definition of valid assessment	3	Self-perceived weakness in definition of valid assessment
Weakness in definition of formal and informal assessment	2	Self-perceived weakness in definitions of formal and informal assessment
Insufficient definition of correct evaluation	1	Self-perceived weakness in definition of correct evaluation

In Table 50, operational definitions of all codes along with direct quotations from the interview records detail the weaknesses of the participants on the basis of definition. It has been revealed that the participants had superficial knowledge about assessment, which was accepted as the most basic definition. Based on the analysis, it can be said that the participants associated the assessment with administering written exams to students in order to discover their learning levels. Although the participants talked about the process and observing students while describing the assessment, for almost all of them, the assessment was to put the students to the exam through pencil and paper exams. The following extracts illustrate the situation:

When I think of assessment, I think of measuring students with written exams. Also, I would like to point out that I did not take many assessment lessons at the university. I have not participated in any in-service training for 8 years. So my knowledge is limited on this subject, I cannot speak confidently. (Semi-structured interview, P3)

4.3.2.3.2. Teachers' Weaknesses in Assessment Types

Some of the questions asked in the interviews aimed to reveal the assessment-type knowledge of the participants. With this sub-theme, the weaknesses of the participants in terms of assessment types were revealed. Considering the self-perceived weaknesses of the participants, it might be said that they had insufficient knowledge about basic assessment types, confused assessment types with formal and informal assessment and had incomplete knowledge about formative assessment, which was a kind of assessment. Table 51 provides detailed information about the results.

Table 51.

Teachers' Weaknesses in Assessment Types

Code	f	Code Description
Perceived weakness in assessment types	7	Weakness in types of assessment. With this code, it was understood that the participants directly stated their weaknesses related to assessment types in general.
Perceived weakness in implementation of formative assessment	5	Weakness regarding formative assessment and its implementation. It was deduced that the participants had superficial knowledge about formative assessment, and that they had weaknesses in terms of knowledge and practice.
Weakness in integration of process in formative assessment	4	Participants explained that formative assessment is a type of assessment related to the process, but they reported that they do not know to incorporate the process in assessment applications.
Associating formative assessment with formal assessment	2	Participants had wrong perceptions with formative assessment and confused formative assessment with formal assessment.

Firstly, when the participants were asked direct questions about assessment types, they stated that they could not remember them at all. After a few clues, the participants answered these questions with only superficial information about formative and summative assessment. Apart from these, they did not provide any information about any other assessment types. Based on this, it can be claimed that the participants had weaknesses in assessment types. Following excerpts explain the situation better:

I couldn't remember the assessment types exactly. I do not know. (Pre-semi-structured interview, P1).

*Aren't assessment types for level, that is, for placement purposes? There is a norm, it was a certain thing, wasn't it? A certain... **I have weaknesses in this regard**, yes. So **I don't know about them**. (Pre semi-structured interview, P2).*

Secondly, this sub-theme shows the weaknesses of the participants regarding knowledge and practice of formative assessment. When the participants were asked questions based on formative assessment, they either said "I don't know" directly to the questions or made incomplete explanations about them with an unsure tone of voice. Based on the statements of the participants, it was clear that they saw formative assessment as an alternative assessment technique, or that they confused it with formal assessment. In addition, teachers expressed that they had weaknesses in using this type of assessment in their practices. The expressions exemplify these ideas:

***What we call formal assessment is what we call formative assessment, right?** (Pre-semi-structured interview, P2).*

*Actually, I have heard of formative assessment **but I do not know to materialize it** (Pre-semi-structured, P3).*

4.3.2.3.3. Teachers' Weaknesses in Purposes of Assessment

During the interviews, the instructors talked about their own assessment purposes. While specifying their weaknesses on this subject, their direct statements, their incomplete or incorrect explanations were determined and then they were analysed. As a result, their weaknesses of the participants in this sub-theme were found out. Teachers' perceived weaknesses considering assessment types revolve around lack of knowledge in assessment purposes, assessing students for punishment or reward, and assessing for providing concrete data to stakeholders. These perceived weaknesses are given in Table 52.

Table 52.

Teachers' Weaknesses in Purposes of Assessment

Code	f	Code Description
Weakness in the purpose of assessment-punishment	5	Participants stated that the purpose of assessing their students is sometimes to punish them.
Lack of knowledge in assessment purpose	2	Participants generally stated that they assess and evaluate in order to measure knowledge level of students. Based on their statements, it was deduced that they do not have sufficient knowledge about the purposes of assessment.
Purpose of assessment for providing managers with concrete data	1	Participants stated that the purpose of assessment is to provide evidence to school administrators about what they are doing.
Purposes of assessment for helping students	1	The participant stated that the purpose of measurement is to help students get high grades.

According to the results, it can be said that the participants had incomplete or incorrect knowledge regarding the purposes of assessment. First of all, the participant instructors clearly stated that they had an intention to punish students in some of their assessment practices. To illustrate, they said that they deliberately asked difficult questions in exams in order to warn some naughty students and that their aim with this application was not to determine whether the students gained the target behaviours, but to warn and punish them harshly. Considering these explanations, it can be said that assessing students aiming at punishing them was a self-perceived weakness of the participants. The following excerpt illustrates this situation:

*In some classes, for example, I applied this in one of the preparatory classes. He wouldn't tell the class so they drove me crazy. I prepared one of the quizzes. I assigned **low scores to easy questions and high scores to difficult questions**. For example, I assigned 30 points to the difficult question. I assigned 10 points to the easy question. **So that students can get a low grade**. I don't do it for revenge, but I rarely **do it for punishment**. Actually, it's for their own good. (Pre-semi-structured, P4).*

Interestingly, one of the participants indicated that s/he assessed just to give high grades to the students. Based on his/her remarks, assessing students for helping them get higher grades or assessing students for providing them a reward can be perceived as one of the weaknesses of participants. Following extract exemplifies the participant's purpose:

*Sometimes I assess for **helping student get a high grade in the exam**. I think, I want the student to get high grades. Sometimes I do assess for this purpose. (Pre-semi-structured, P1).*

Furthermore, it has been determined that some teachers fulfil the expectations of the institution, even though they do not support this, just because the institution they work for requires them to evaluate the students by using the written exam technique three times during the semester and to keep these written exams as exam documents. This result can also be described as a perceived weakness of teachers in terms of assessment purpose.

4.3.2.4. Teachers' Weaknesses about Assessment Results

When the instructors were asked questions about the assessment results, their answers were grouped under three sub-themes indicating their weaknesses related to this subject. They are teachers' weaknesses in interpreting assessment results, teachers' other weaknesses related to assessment results and teachers' weaknesses in using assessment results. While the teachers' weaknesses in interpreting assessment results was the most mentioned sub-theme, it was followed by teachers' other weaknesses related to assessment results and teachers' weaknesses in using assessment results respectively.

4.3.2.4.1. Teachers' Weaknesses in Interpreting Assessment Results

When the participants were asked questions about the interpretation of the assessment results, it was revealed that they had weaknesses related to this subject. Participating instructors openly stated that they somehow interpreted the results, but that they did not know any statistical calculations for this and naturally they did not use them in their practices. It was also been observed that the teachers confused the interpretation of the data obtained as a result of the assessment with giving feedback to the students. Table 53 presents the codes that make up this sub-theme with their frequencies and descriptions:

Table 53.

Teachers' Weaknesses in Interpreting Assessment Results

Code	f	Code Description
Weakness in interpreting assessment results	19	With this code, it was deduced that the participants felt weakness in interpreting the results that they obtained as a result of assessment practices.
Insufficient knowledge of avoiding misinterpretation of results	5	Self-perceived weakness in avoiding the misinterpretation of results. Here, it was revealed that the participants felt inadequate to avoid misinterpreting the assessment results.
Nature of interpretation of assessment results by students	3	It was noticed that the participants felt deficiencies in perceiving the way in which the assessment results obtained by the participants were interpreted by the students.
Nature of interpretation of assessment results by authorities	2	It was noticed that the participants felt deficiencies in perceiving the way in which the assessment results obtained by the participants were interpreted by the administrators.
Confusing result interpretation with giving feedback	1	It was deduced that the interpretation of assessment results was confused with the act of giving feedback to the students by looking at the results after the assessment process.

From the answers of the participants about the interpretation of the assessment results, it was revealed that teachers had some conceptual confusions about this subject and that they had weaknesses about how this interpretation should be materialized as a teacher. Moreover, as a result of the analysis, we found that the participants had also weaknesses considering the knowledge in nature of both students' and administrators' interpretation of assessment results.

Looking at the statements of the participants, it was seen that their answers about the interpretation of the assessment results were far beyond statistical calculations. It was found that while the teachers were interpreting the results, they only made comments about the general success level of the students by looking at the results without making any mathematical calculations. This finding is exemplified:

*That is, I interpret the assessment results on a class basis. "Oh, this class has learned very well, this class has not learned well" I can say by interpreting assessment results. I don't make interpretations on the basis of individual students, but there are some such students that attract the attention of the teachers. By looking the exam results, we can see that they got 40. Then I say "Oh, I wonder who got 40?" You look at his name saying who is this student? This is how I interpret the assessment results. **Apart from these, I do not know how to make interpretation** (Pre-semi-structured interview, P3).*

Later, from the answer given by the participant, it was revealed that they confused the interpretation of assessment results with giving feedback to the students. When the instructor talked about his/her practice of interpreting the results, s/he is actually talked about the act of giving individual feedback to the students. The following excerpt illustrates the participant's misconception regarding the interpretation of the assessment results:

*I usually explain to students one by one. I contact the student by saying, "Roughly, you need to study a little more or you are studying well." **But it is not very clear, these are implicit sentences. This is how I comment.** (Pre-semi-structured interview, P1).*

From the participant's statement, it might be claimed that his/her weakness is that s/he perceives giving feedback to students as interpreting the assessment results.

4.3.2.4.2. Teachers' Other Weaknesses Related to Assessment Results

Teachers' answers focusing on assessment results revealed some of their weaknesses about the issue and they were categorized under the heading of teachers' other weaknesses in assessment results. Their self-perceived weaknesses in this sub-theme were about ensuring the validity, reliability and consistency of the assessment results, preventing inclusion of

measurement errors in assessment results and sharing the assessment results with stakeholders. Table 54 presents the participants' weaknesses in this sub-theme in detail:

Table 54.

Teachers' Other Weaknesses Related to Assessment Results

Code	f	Code Description
Weakness in ensuring consistency of assessment results	8	Perceived weakness in consistency. Here, it was deduced that the participants have incomplete knowledge about ensuring consistency in the assessment results.
incomplete knowledge in sharing assessment results	5	Perceived weakness in sharing assessment results. It has been revealed that the participants have weaknesses related to the way of explaining the assessment results.
Weakness in enhancing reliability of assessment results	3	Perceived weakness in enhancing the reliability of assessment results. It was understood from the statements of the participants that they do not have enough knowledge to ensure the reliability of the assessment results.
Weakness in enhancing validity in assessment results	2	Self-perceived weakness in the validity of assessment results. It was understood from the statements of the participants that they do not have enough knowledge to ensure the validity of the assessment results.
Weakness in preventing measurement error in assessment results	2	Self-perceived weakness in preventing measurement errors in assessment results. From the statements of the participants, it was understood that they had weaknesses in preventing inclusion of measurement errors into assessment results.

Weakness in ensuring consistency of assessment results was determined as the most frequently expressed code by the participants. This code gives the message that the participants had insufficient knowledge namely weakness regarding consistency in the assessment results. By consistency, the participants referred that the grades that the students receive from the exams throughout the process should be similar to the grades expected from them. The following quotes illustrate this situation:

*I also consider the student's performance in the classroom. I can also compare the student's grades with previous exams and reach a conclusion. **If a child who performs poorly in the class scores high in the exam, there is an inconsistency.** This is what I mean by consistency. (Pre-semi-structured interview, P2)*

In addition, the weaknesses of the participants in ensuring reliability of the assessment results were also revealed. They just stated that they were very careful while giving grades and gave everyone the grade they deserved. Apart from this, they stated that they did not do anything else to ensure the reliability of the assessment results. This emerging result suggests that it aligns with the teachers' strengths concerning ensuring objectivity in grading and

ethical considerations. The following excerpt illustrates how the participants ensured the reliability of the assessment results.

*I ensure the reliability of the assessment results as follows: **When I grade correctly and give everyone the grade they deserve appropriately, then I ensure reliability.***

(Pre-semi-structured interview, P3)

In order to prevent inclusion of measurement error into assessment results, the participants remarked that they only prevented students from cheating which was also related to teachers' perceived strengths regarding administering exams. The fact that they only prevented cheating and that they did not know and apply anything else about the issue showed that they had weaknesses in preventing the inclusion of errors in the assessment results. Participant 1 exemplifies how s/he achieves preventing measurement error in his/her assessment results:

*I try to get my results as error-free as possible in order to avoid including errors in the assessment results. **The most basic way I apply this is to prevent students from cheating during exams.** When they can't cheat, everyone gets the grade they deserve. The results that emerge in this way are naturally error-free. Other than that, I don't pay attention to anything else. I don't know if there is, but this is the most effective, so I don't think there is any need for an extra application. So I can only say that I pay attention to it. (Pre-semi-structured interview, P1)*

4.3.2.4.3. Teachers' Weaknesses in Using Assessment Results

Through this sub-theme, teachers' weaknesses in utilising assessment results are provided. Teachers were found to use assessment results only in determining students' learning level, making decisions on their success, and evaluating teaching effectiveness. Apart from these, they reported not using the results. In addition to these, teachers' weaknesses regarding use of assessment results were found to be related to use of assessment result for planning the instruction, making adaptations in the curriculum as well. Table 55 details the sub-theme:

Table 55.

Teachers' Weaknesses in Using Assessment Results

Code	f	Code Description
Weakness in using assessment results	5	Self-perceived assessment results in the usage of assessment results. The participants mentioned that the knowledge of the usage area of the assessment results is limited.
Weakness in how to make adaptations in curriculum based on results	4	Self-perceived assessment results in making adaptations in curriculum depending on assessment results. The participants stated that they did not have enough knowledge to make necessary changes in the curriculum based on the assessment results.
Weakness in using assessment result in changing teaching plans	2	Self-perceived weakness in using assessment results in changing teaching plans. The participants stated that they did not have enough knowledge to make necessary changes in the teaching plan based on assessment results.

First of all, general questions about the use of assessment results were asked to the instructors. Their weaknesses were revealed after their strengths were pointed out from their answers. It was concluded from the answers that the participants had some weaknesses in using assessment results. In other words, as it was same in teachers' perceived strengths related to the current issue, teachers were found to use the assessment results to determine the learning levels of the students, to decide on their success, and to evaluate the effectiveness of their own teaching. However, teachers did not mention using the assessment results in the development of the institution, changing the assessment tool or developing it, improving teaching plans and the curriculum. Based on these results, we can say that teachers had insufficient knowledge regarding this issue. The following excerpt illustrates these statements:

*I use the assessment results to **determine whether the student has learned the target subjects or not**. We make these inferences by looking at the results and collectively decide that the student should pass and fail. Apart from these, I do not use results.*
(Pre-semi-structured interview, P3)

In addition, weaknesses were found in making changes in participants' own teaching plans by considering assessment results. By regarding the results, the participants stated that they somehow made adaptations in their teaching, but they did not have information about doing it systematically.

Similarly, teachers also stated that they had weaknesses in making changes in the curriculum. By looking at the assessment results, they explained that they somehow made changes in the curriculum, as in making changes in the teaching plans, but they had incomplete knowledge on how to do this systematically. The following excerpt explicates the situation:

*Based on the assessment results, the curriculum should be changed. I try to do this as much as possible, but within my means, of course. **I don't have an idea about how to do this.** (Pre-semi-structured interview, P1)*

*Of course, **some changes should be made in the curriculum** according to the weaknesses of the student by looking at the measurement results. But I don't do this very often because we have to complete the curriculum. We have 3 textbooks. All books must be completed. While these responsibilities exist, **how can changes be made in the curriculum by looking at the assessment results? I may fail to both achieve the goals and make changes in the curriculum based on the results simultaneously** (Pre-semi-structured interview, P2)*

4.3.2.5. Teachers' Weaknesses Related to the Assessment Technique

During the investigation, teachers' weaknesses related to assessment techniques were also revealed. On the basis of the analysis, three broad sub-themes were identified. The most frequently noted sub-theme was the teachers' weaknesses in knowledge and implementation of assessment technique. Then, it was followed by the teachers' weaknesses in criteria for technique selection, and the teachers' weaknesses in the purpose of technique usage respectively. Sub-themes belonging to this theme were explained in detail in the following sections.

4.3.2.5.1. Teachers' Weaknesses in Knowledge and Implementation of Assessment Technique

In the interviews, questions were asked to the teachers to determine their weaknesses in knowledge and practice of assessment techniques. While answering the questions, teachers stated that their knowledge of assessment techniques was weak in general. Furthermore, they stated that they did not have sufficient knowledge about the types of assessment techniques, and they did not know which technique they should use for what purpose and how. Therefore, they reported preferring only the techniques which were totally known by them. Table 56 presents detailed information about this sub-theme

Table 56.

Teachers' Weaknesses in Knowledge and Implementation of Assessment Technique

Code	f	Code Description
Lack of knowledge of kinds of assessment techniques	15	This code gives the information that the participants have heard of various assessment methods, and know some of them by name, but generally have weaknesses in the knowledge of assessment techniques.
Weakness in general knowledge in assessment techniques	10	Self-perceived weakness in having general knowledge about assessment techniques.
Incomplete knowledge considering pros and cons of techniques	5	This code informs that participants do not have enough information about the strengths and weaknesses of various assessment and evaluation methods.
Using an assessment method without doing research	4	Participants state that they choose an assessment tool or method by heart without doing any research while choosing the assessment method for their applications.
Using only assessment method whose application is known	3	Participants only use techniques they know, they never use what they do not know, and thus the methods they use are limited.

During the interviews, nearly all of the participants stated that they did not have enough knowledge about the kinds of assessment techniques. They also added that although they were aware of the existence of various techniques, they could not use them in their own assessment practices since they did not know them sufficiently. This result showed that the reason for teachers to use a single assessment technique in their practices was their lack of knowledge. Based on this result, we can say that if teachers have sufficient knowledge about the kinds of assessment techniques, they will diversify their practices accordingly.

*In other words, as I mentioned before, there are many assessment techniques. There are traditional techniques, there are alternatives. **This topic has a wide range and I do not have detailed knowledge about it.** Therefore, I cannot say that I feel competent. (Pre-semi-structured interviews, P1)*

*Frankly, we learned about these subjects when I was an undergraduate student and I learned a lot, but time passed and I inevitably forgot them. **Since I have been using the same type of assessment technique at this school, that's why I know a limited number of assessment techniques.** (Pre-semi-structured interviews, P2).*

In addition, the participants stated that they were not informed about advantages or disadvantages of any assessment technique. In other words, they provided the information that they did not have any knowledge about which assessment technique to use for which

purposes. Therefore, they reported using the techniques that they had been experienced and accustomed to using them for a long time without doing any research. The excerpts explicate these ideas:

It would be an extra burden for me to do research beforehand for each technique.

Applying a new technique that I am not fully familiar with, namely poses a risk.

Problems may arise when I apply an assessment technique that I am not fully knowledgeable about. Therefore, since I do not do research, I only use techniques that I am knowledgeable about. (Pre-semi-structured interviews, P1)

I prefer assessment techniques that we are all used to and that we have been using for a long time. (Pre-semi-structured interviews, P3)

Based on these results, we can say that the weaknesses of teachers' knowledge of assessment techniques limit their classroom practices. In addition, it has been determined that teachers prefer using the techniques they have been familiar with since their student years, without doing any research and not even knowing why they use the technique.

4.3.2.5.2. Teachers' Weaknesses in Criteria for Technique Selection

As a result of the analysis, it was also revealed that another weakness of the participants was related to the criteria that should be considered when choosing any assessment tool or technique. When the participant instructors were asked questions related to this issue, it was found out that they considered only practicality and content validity and that they did not know any other criteria apart from them. Table 57 gives detailed information about the results related to this current sub-theme.

Table 57.

Teachers' Weaknesses in Criteria for Technique Selection

Code	f	Code Description
Weakness in knowledge of criteria for selection of method	11	Participants stated that they had weaknesses in terms of what they should consider when choosing a valid and reliable assessment tool or method, in other words, which criteria they should know.
Weakness of choosing assessment method relevant to objectives	8	Self-perceived weakness in selecting assessment technique relevant to objectives. Here, the participants stated that they did not feel themselves competent enough when choosing an assessment tool or technique related to their objectives.
Influence of previous learning experiences in selection of technique	1	It is understood from the statements that experiences of the participant as a student are effective in determining the assessment technique that the participant will use as a teacher. Self-perceived weakness in selection of assessment method, choosing assessment method based on previous learning experiences.

According to the results of data analysis, the participants stated that they did not have enough knowledge about what to consider when choosing an assessment technique or tool which was suitable for their assessment purposes. They paid attention to an assessment tool or technique with content validity that only assesses whether the subjects taught to students had been learnt or not and did not contain questions related to the subject that had not been taught. In addition, they paid attention to the fact that a technique or an assessment tool should be practical for both students and teachers. Apart from these, they said that they did not know any criteria to be known. The following excerpts clarify this situation:

Frankly, I don't know what to consider when choosing or adapting an existing assessment tool or when developing a new one. I download it directly from the internet. I just take into consideration whether the tool I'm using is in line with the topics or not. (Pre-semi -structured interview P1)

First of all, I choose techniques that I am familiar with, that I can apply in a practical way, and that the students are also familiar with. I do not choose a technique or tool that may create difficulties for the students and for me. As I said, my assessment tool should measure what I have taught, not too much. So it should be valid. (Pre-semi structured interview P4)

In addition, participants preferred the method(s) administered to them as they were students when choosing an assessment technique or tool for their practices. For this reason, it might be said that assessment practices of the participants in their previous learning experiences were found to be effective in determining the technique or assessment tool in their own practices as teachers. Participant 2's statements clarify the situation:

*When we were in high school, when we were undergraduate students, the methods I apply now are applied to us. **Our past learning experiences affect us. I haven't done much research.** (Pre-semi-structured interview P2)*

In fact, these results were also revealed when determining the strengths of teachers related to the subject. To go into detail, teachers' self-perceived strengths regarding criteria used in selection of assessment techniques were found as practicality and validity and then we can say that teachers want to avoid problems while applying and grading the exams and to complete these processes easily.

4.3.2.5.3. Teachers' Weaknesses in the Purpose of Technique Usage

In this theme, lastly, the sub-theme that shows the weaknesses of the participants related to the purpose of using assessment techniques is included. During the interviews, most of the participants stated that their aim was to determine the learning level of the students, and they did not have a pivotal aim apart from this. The following table describes the results:

Table 58.

Teachers' Weaknesses in the Purpose of Technique Usage

Code	f	Code Description
Insufficient knowledge in purposes of using techniques	4	This code informs that the participants' purposes of assessment techniques are limited.
Use of assessment techniques to get a concrete evidence	2	This code informs that there is a mistake in the purpose of using assessment techniques of the participants and that they also have a weakness in the purpose of using the assessment technique.

As stated before, the participants determined that the purpose of using assessment techniques was to assess the knowledge level of students. Apart from this, they did not know for what purposes assessment techniques could be used. It was always found out that the aim of the teachers was to determine the learning level of the students. Looking at this result, we can say once again that teachers see assessment as only determining the learning levels of students as a result of teaching. In addition, the participants indicated that they used assessment techniques to meet the assessment expectations of the institution they worked for and to provide a document that officially proves that assessment and evaluation have been made. Consequently, participants might be said to have weaknesses considering the purposes of assessment technique/tool usage.

4.3.2.6. Teachers' Weaknesses in Grading and Scoring

Teachers' weaknesses regarding assessment were found to be related to scoring and grading as well. These perceived weaknesses were grouped under two sub-themes. Teachers' weaknesses in grading, the most noted one, was followed, respectively, by teachers' weaknesses in scoring. Each sub-theme was explained separately below.

4.3.2.6.1. Teachers' Weaknesses in Grading

With this sub-theme, information about weaknesses of the participants regarding grading is provided. It revolves around the issues of validity and reliability in grading, grading purposes, and how to grade. Table 59 presents the codes included in this sub-theme:

Table 59.

Teachers' Weakness in Grading

Code	f	Code Description
Perceived weakness in enhancing reliability in grading	10	It includes the weaknesses of the participants in terms of ensuring reliability in grading.
Perceived weakness in how to grade	7	This code gives information about the self-perceived weaknesses of the participants about how they should do about grading in their assessment practices.
Perceived weakness in enhancing validity in grading	6	It includes the weaknesses of the participants in terms of ensuring validity in grading.
Perceived weakness of why to grade	4	Based on the statements obtained, it was deduced that the participants had insufficient knowledge about why to grade.

The most frequently stated code in this sub-theme is ensuring reliability in the grading. As a result of the analysis of participants' statements teachers were found as having weaknesses concerning this issue. While answering the questions about grading and reliability, the teachers stated that were objective and only gave the grades to those who really deserved them along with using an answer key. They had clearly indicated that they did not know anything else related to this subject apart from the practices they mentioned and that this was their self-perceived weakness. The excerpts exemplify this result:

For example, I don't give grades on my own during grading. I give each student the grade they deserve. I don't give extra or low grades. I don't know of any other ways to ensure reliability in grading. (Pre-semi-structured interviews, P1).

Reliability in grading... In other words, I ensure reliability in grading by preventing students from cheating in the exam. In this way, the results are also reliable. I act objectively when grading. My results will be reliable. Other than that, I don't know anything else. (Pre-semi-structured interviews, P1).

Moreover, the participants were asked questions about why they were grading. Based on their answers to the questions, it can be inferred that the participants only assessed knowledge level of students and graded to obtain numerical data that would later be used to decide whether the students should pass the class, or not. This emerging result was found to be in line

with teachers' perceived purpose of assessment. Teachers claiming to conduct assessment to shed light on students' learning levels and make decisions on their success were also found to have similar purposes while they were giving graders. The excerpt exemplifies the result:

Why should I grade? hmmm. I honestly can't say anything. I give grades to students' performances or their written responses. Other than that, I don't know if there is a special purpose. (Pre-semi-structured interview, P1)

The lecturers, who only talked about being objective concerning the issue, did not touch upon an extra point to ensure reliability in grading. From the statements of the teachers, it was revealed that teachers did not have enough knowledge to increase the reliability in grading.

4.3.2.6.2. Teachers' Weaknesses in Scoring

Participants also answered questions about scoring during the interviews. It should be noted that weaknesses of the participants considering definitions of both scoring and grading, were also revealed as a result of the analysis. They were reported under different sub-themes in previous parts. In this sub-theme, the participants perceived weaknesses in scoring were appeared in how to score, and how to ensure reliability and validity in scoring. Table 60 presents the weaknesses of the participants concerning this topic:

Table 60.

Teachers' Weaknesses in Scoring

Code	f	Code Description
Perceived weakness in how to score.	10	Based on the statements of the participants, it can be said that the participants performed this process without relying on any information while scoring.
Weakness in scoring in general	7	With this code, it was deduced from the statements of the participants that they had some weaknesses in the general scoring process.
Weakness in ensuring validity in scoring	4	Self-perceived weaknesses in enhancing validity in scoring.
Weakness in enhancing reliability in scoring	3	Self-perceived weaknesses in enhancing reliability in scoring.

It is apparent from the Table 60 that teachers behaved in favour of students while scoring. In other words, teachers stated that they deliberately assigned high scores to easy questions so that students get higher grades in the exam. This result shows the weakness of the teachers in terms of scoring as well as a result that contradicts the previous findings emerged as a result of the analysis. To illustrate, although the teachers said that they gave everyone the grade they deserved during the grading and that they tried to reveal the real competency level of the students with the exam results, the fact that the teachers behaved in a

way to help students get higher grades in scoring shows this contradiction. Moreover, teachers also stated that they used scoring as a reward or punishment tool, that is, they assigned higher scores to the easy questions to the class they liked, while they assigned higher scores to the difficult questions of the class they disliked. The statements of the participants showing teachers' weaknesses in scoring exemplify this situation:

Now I take the student as the basis. I score according to the level of the class. I give a higher points to the very easy part so that students can answer correctly and get a high score. (Pre-semi structured interviews, P4).

Sometimes I assign low scores to the hard questions of my favourite class and high scores to the easy questions so that the students can get high scores, so that I can give them a helping hand. (Pre-semi-structured interviews, P3)

In addition to these, it was concluded that participants had weaknesses in terms of ensuring reliability in scoring. They indicated that they did not know what to do in order to assign scores reliably and to be able to assess information or behaviour by avoiding from inclusion of measurement error. Moreover, they stated that they did not do anything special to ensure reliability in scoring and that this was the first time that they had ever heard that reliability could be considered in scoring. The following excerpt explains the situation:

So, for example, I see myself as competent for which subject and how much scoring will be given in the exam, but I don't know if this is validity or reliability. (Pre-semi-structured interviews, P3)

4.3.2.7. Teachers' Additional Weaknesses in Assessment

As a result of the analysis of the interview records, which were organized to determine the weaknesses of the participants in the assessment, a theme called teachers' additional weaknesses in assessment emerged. In this theme, teachers' weaknesses about ethics in assessment and teachers' other assessment-related. These sub-themes were explained one by one in the following parts.

4.3.2.7.1. Teachers' Weaknesses about Ethics in Assessment

After participants had answered questions about ethical issues in assessment in the interviews, it was found out that they had some weaknesses regarding this topic. Instructors associated ethical issues only with grading exams. In other words, according to the participants, behaving ethically in assessment means to be fair, objective in grading and to

treat everyone equally. Apart from these, the participants did not make any additions related to ethical considerations in assessment. Table 61 details this sub-theme:

Table 61.

Teachers' Weaknesses about Ethics in Assessment

Code	f	Code Description
Perceived weakness in ethical considerations in assessment	12	Self-perceived weakness in ethical considerations. Here, it can be said that the participants feel lack of knowledge about how to consider ethical issues in assessment practices, in other words, what they should know while applying ethical issues in assessment.
Weakness in implementation of ethical considerations in assessment	8	Self-perceived weakness in implementation of ethical issues in assessment practices. It can be concluded that since the participants do not know the ethical issues in assessment, they cannot apply this issue in their practices.
Weakness in coping with unethical behaviours in assessment	5	Self-perceived weakness in coping with unethical behaviours in assessment practices. This code gives the message of the lack of knowledge that will be effective in coping with unethical behaviours in assessment.
Wrong attribution of unethical behaviours in assessment	1	In fact, it can be said that the participants do not even know exactly which behavior is unethical in assessment.
Insufficient knowledge about ethics in grading	1	Although it is clear that the participants have partial knowledge of ethics in grading, it is clear that they lack of knowledge about this subject.

Looking at the results, it is seen that the most frequently noted sub-theme is perceived weaknesses in ethical considerations. Participants stated that they only paid attention to ethical issues in grading, but they reported to have insufficient knowledge in considering ethical issues in other areas related to assessment.

*When ethical issues in assessment are mentioned, I think of **treating everyone equally, being fair and being objective in grading**. What else could it be? If there is, I can't remember right now. (Pre-semi-structured interview, P2).*

In addition, when the participants were asked about coping with unethical behaviors that emerged in their assessment practices, the participants stated that they had not encountered such a situation so far and that they did not know how to behave in this situation. Based on their statements, it can be said that the participants had weaknesses related to the subject. Direct quotations from interviews exemplify the situation:

I would continue to do my job by ignoring that unethical behavior. Frankly, I have not encountered such a situation, but if I did, I would act as I mentioned. I don't know how to behave otherwise. (Post-semi-structured interview, P1).

4.3.2.7.2. Teachers' Other Assessment Related Weaknesses

With this sub-theme, information about weaknesses of the participants regarding administering exams, avoiding from measurement error and informing students about grading is given. Table 62 below gives detailed information about this sub-theme:

Table 62.

Teachers' Other Assessment Related Weaknesses

Code	f	Code Description
Limited knowledge in administering exams	8	Weakness in considerations of administering exams. It was deduced that the participants did not pay attention to anything other than preventing cheating while administering exams to the students, and that participants might have weakness in knowledge of this issue.
Inadequate knowledge about preventing measurement error	7	Weakness in considerations of preventing the inclusion of measurement error into assessment practices. It was understood that that the participants did not pay attention to anything other than preventing cheating and giving clear instructions to students while administering exams, and that participants might have weakness in knowledge of how to prevent measurement error.
Lack of knowledge about informing students' about assessment	4	It is thought that the participants only talk about how many exams will be applied to the students during academic year along with their impact rates. Except for these, It is also acknowledged that the participants do not give detailed explanations to the students about assessment philosophy and they do not know why they should do it.

When Table 62 is examined, it is seen that the participants had incomplete knowledge about the principles of applying exams. They stated that they paid attention to prevent cheating and to arrange the physical environment of the classroom for the exam while administering the exams. Apart from these, teachers did not make any statements on this issue. On the basis of the results, we can say that the participants had a superficial knowledge of what should be considered while administering the exams.

Secondly, we can say that teachers also had weaknesses in preventing measurement errors. They indicated that while trying to avoid measurement errors, they only prevented cheating, they tried to give clear statements in exams, and they were objective in grading. As it can be understood from the explanations of the instructors, it can be said that the participants tried to prevent the measurement error only in the exam application and grading. Apart from these, they did not give any explanation about the possible sources or types of errors and what can be done at other stages of assessment to prevent inclusion of measurement error. The following excerpts illustrate this result:

I would try to make my instructions clearer so that the student could understand the questions correctly. (Post-semi-structured interview, P1).

In other words, because I associate measurement error with cheating more or because I associate it with objectivity while grading the teacher's exam results, I do not cheat in exams, and I act objectively while giving grades. In this way, I think to the students that mistakes are not mixed in the results. (Pre-semi-structured interview, P3)

The detailed explanation of the last code of this sub-theme, the lack of knowledge about informing students about assessment, can be found in Table 62 above.

4.3.2.8. Teachers' Weaknesses Regarding Assessment Principles

Some of the questions asked to the participants were about the principles of assessment. Teachers talked about validity and reliability superficially. Apart from these, they did not mention any other principles. From the teachers' answers, three sub-themes emerged such as teachers' weaknesses in validity, teachers' weaknesses in reliability, teachers' weaknesses in influence of assessment were created. Since teachers did not mention other language assessment principles such as washback, authenticity, practicality, their weaknesses in those subjects could not be included in these sub-themes. However, it was accepted that the teachers also had weaknesses considering other principles. The aforementioned emerging sub-themes are explained in detail in the following sections.

4.3.2.8.1. Teachers' Weaknesses in Validity

The most obvious weakness of the participants related to validity is what validity means as a concept, its types and how to increase validity in assessment applications. Table 63 shows the detailed information for this sub-theme:

Table 63.

Teachers' Weaknesses in Validity

Code	f	Code Description
Insufficient knowledge about validity	12	Weakness in validity. In general, participants do not know anything about validity other than content validity. The importance of validity, its types, etc. were understood from the statements that the participants did not know.
Weakness inability to develop valid assessment tool	4	It was also revealed that the participants lacked in developing a valid measurement tool and creating a valid measurement process in relation to validity.
Perceived weakness in ensuring validity in assessment	1	Self-perceived weakness in how to enhance validity in assessment practices.

When the Table 63 is examined, it is seen that the participants had a general weakness in knowledge about validity. It has been revealed that when the word validity was mentioned,

the instructors only thought of content validity and did not talk about other types of validity. Some participants even stated that they did not know anything about validity other than content validity. Based on these, it can be deduced that the participants did not have enough information about validity, which is a principle, and that they had weaknesses related to it. The following quote exemplifies this situation:

*Validity, that is, as far as I remember, from KPSS... **When I say validity, content validity comes to my mind directly.** The content, that is, the exam should be compatible with the content, **it should be appropriate with the questions I ask.** In other words, the topics covered in the course and the questions asked in the exam are the same. **The exam should not include extra topics.** I can't say anything else about it. (Pre semi-structured interview – P3)*

Regarding validity, the participants still felt inadequate to develop a valid assessment tool for their own assessment practices. While answering the question, the instructors answered only on the basis of written exams as an assessment tool, and they stated that they only provided content validity for the sake of validity while preparing these tools and did not pay attention to anything other than that. This result was in line with the results revealed by previous code. In other words, since the participants' knowledge of validity was limited, they also had weaknesses in preparing a valid assessment tool.

*Although I remember some things about the subject in my head, **this information is not always complete.** I am not sure if what I remember is correct. Paper-pencil exams that I regularly prepare and apply are my assessment tool and **I ensure their validity with content validity.** Other than that, **I have nothing to pay attention to.** (Pre semi-structured interview, P1).*

4.3.2.8.2. Teachers' Weaknesses in Reliability

Based on the statements of the participants, their reliability-related weaknesses were classified within three codes. These are ensuring reliability in assessment, enhancing reliability while administering exams, and selecting a reliable assessment tool or technique. Table 64 below details this sub-theme:

Table 64.

Teachers' Weaknesses in Reliability

Code	f	Code Description
Weakness in enhancing reliability in administering tests	5	Perceived weakness in ensuring reliability during the application of exams
Perceived weakness in ensuring reliability in assessment	3	Perceived weakness in enhancing reliability in general assessment practices
Weakness in selecting a reliable assessment technique	2	Still, it is related to reliability and participants talk about self-perceived weaknesses in choosing reliable assessment tools/techniques.

First, it was revealed that the participants had weaknesses in enhancing reliability while administering the exams. Instructors stated that they only tried to ensure reliability by preventing cheating and by giving clear instructions to students while administering the exams. Apart from these, the participants did not mention any measures or practices. The following excerpts illustrate this situation:

*I guess I don't provide **reliability while applying the exams** (Pre-semi-structured interview, P4).*

*So, I try **not to make students cheat**. I make groups A-B in the questions I ask. This is what I mean by **reliability**. **This way I ensure reliability**. (Pre-semi-structured interview, P3).*

Later, when the participants asked questions about how they ensured reliability in general assessment practices, the instructors talked about their exam practices again. In fact, this situation was in parallel with the result reported in the previous part of the study. Since the participants mostly associated assessment with taking an exam even a written exam, when they were asked to talk about general assessment, they involuntarily talked about applying the exams again. Naturally, they referred to preventing cheating and being objective in giving grades. By considering them, participants believed that they could ensure reliability in their assessment practices. The following excerpt exemplifies the situation:

*I have some weaknesses in this regard, but still I am not so bad. But I need detailed information. In fact, I am **somewhat aware of the concepts and practices we have talked about so far**. **But I do not have in-depth or complete knowledge about them**. Everything is superficial... When it comes to **reliability in assessment**, students should not cheat, everyone should get the grade they deserve. (Pre-semi-structured interview, P2).*

4.3.2.8.3. Teachers' Weaknesses in Influence of Assessment

With this sub-theme, the participants' weaknesses were found about the effects of assessment and evaluation practices on teachers' teaching and students' learning. Table 65 presents this emerging sub-theme:

Table 65.

Teachers' Weaknesses in Influence of Assessment

Code	f	Code Description
Incomplete knowledge about influence of assessment, washback	6	Self-perceived weakness in knowledge of how assessment practices influence learning.
Influence of assessment on teachers' practices	2	Self-perceived weakness in knowledge of how assessment practices influence learning.

Participating instructors stated that the assessment practices had an emotional effect on the students. In other words, according to the participants, students took the lesson and the teacher serious thanks to the exams. While answering these questions, they did not mention any other effect of assessment practices on students.

In addition, the participants could not answer the questions about what kind of effects the assessment practices would have on the student or the teacher. Also teachers reported not to know how the assessment practices shape and influence the teaching and learning. Considering the two different issues, it can be concluded that EFL instructors had some weaknesses in how assessment practices influence how teachers teach and how students learn. In order to better understand the results revealed through these codes, excerpt exemplifying them can be accessed below:

*I think the effect of assessment and evaluation is of course on students. Thanks to the exams, **the student does not misbehave in the lesson and takes both the lesson and the teacher seriously.** Thanks to the exam, the student is afraid and does not disturb the class order. (Pre-semi-structured interview, P4)*

4.4. What is the Perceived Development of EFL Instructors Considering Language Assessment, if any?

After teachers' language assessment-related weaknesses had been identified, a series of RRD sessions were designed and held. Afterwards, we aimed to determine whether there was a development related to language assessment in the cognitions of the teachers as they perceived. In order to attain this goal, data were collected through various instruments such as

two separate individual pre-semi-structured interviews with each of the instructors. The first interview included similar questions which were asked to teachers during the pre RRD sessions and it is called as last pre-semi-structured interview. The second interview which was held with each participants in the last phase of the study included not only similar questions but also answers given by the teachers in the pre RRD sessions. This instrument is also called comparison pre-semi-structured interview. With this interview, we aimed to make participants remember their answers before the RRD sessions and to learn whether they would want to add something to their already given answers. Additionally, non-participant classroom observation with field notes, last post-semi-structured interviews, KWL charts, participant reflective reports and two different researcher reflective reports one of which was kept at the end of the study considering the overall development of each participant, the other one was kept immediately after each reflective reading and discussion session. Data derived from these sources were transcribed verbatim and analysed based on inductive content analysis. Throughout the analysis, Maxqda 2020 was utilized. As result of the analysis, nine broad themes emerged considering participants' development in their cognitions about language assessment. The themes are listed according to their frequency of citation in the Table 66 below.

Table 66.

Teachers' Developments Considering Language Assessment

Themes	Sub-themes
Teachers' developments concerning assessment techniques	<ul style="list-style-type: none"> - Teachers' developments regarding performance assessment and rubric - Teachers' developments in knowledge of alternatives - Teachers' developments in specific alternative techniques - Teachers' developments in what to consider while developing, choosing or adapting a technique - Teachers' developments about assessment technique - Teachers' developments in purpose of using assessment techniques
Teachers' developments in writing test items and test construction	<ul style="list-style-type: none"> - Teachers' developments in writing kinds of test items - Teachers' developments in content of a test in test preparation - Teachers' developments in what to consider while writing test items - Teachers' additional developments regarding test preparation - Teachers' developments considering assessment principles in test preparation
Teachers' developments regarding assessment results	<ul style="list-style-type: none"> - Teachers' developments in use of assessment results - Teachers' developments in sharing assessment results - Teachers' developments in interpreting assessment results - Teachers' developments in making decisions based on assessment results - Teachers' developments in reflection of assessment results
Teachers' additional developments regarding	<ul style="list-style-type: none"> - Teachers' developments in ethics - Teachers' developments in administering exams - Teachers' developments in characteristics for quality assessment - Teachers' developments in what to consider to increase the quality of

assessment	assessment
Teachers' developments in introductory issues about assessment	- Teachers' developments in relationships in the issues regarding assessment - Teachers' developed perceptions and definitions of assessment - Teachers' developments regarding assessment type - Teachers' developments of assessment purposes
Teachers' developments in assessment principles	- Teachers' developments in washback - Teachers' developments in validity - Teachers' developments in consistency
Teachers' developments in grading and scoring	- Teachers' developments in grading in general - Teachers' developments of reliability in grading - Teachers' additional developments in grading and scoring - Teachers' developments of ethics in grading
Teachers' developments in reliability	- Teachers' developments in ensuring reliability in administering exams - Teachers' developments in factors influencing reliability in assessment
Teachers' developments in measurement error	- Teachers' developments in knowledge of measurement error - Teachers' developments in how to prevent measurement error

In the following sections, the emerging themes are presented in detail separately in order for better understanding of research findings.

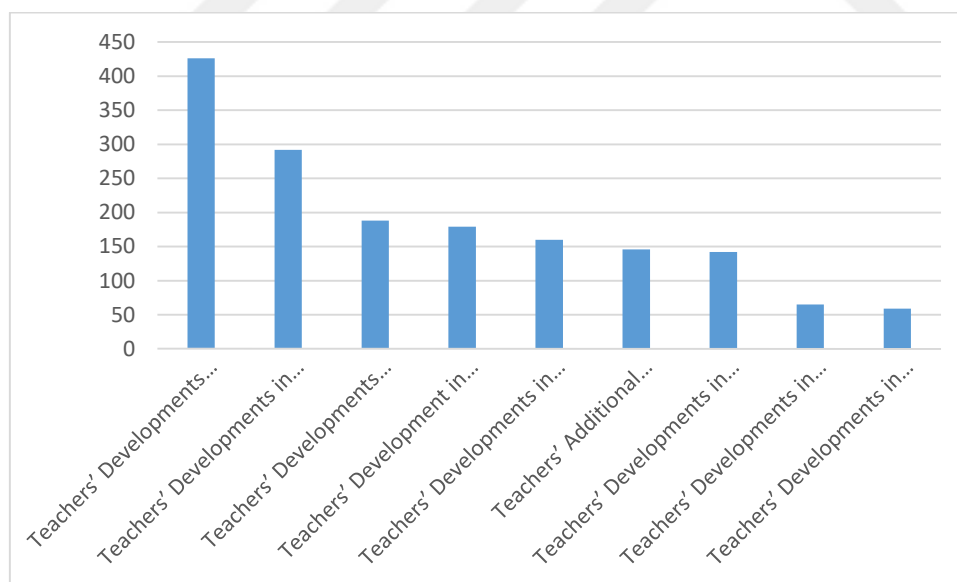


Figure 6. *Teachers' Developments Considering Language Assessment*

Teachers participated in ten RRD sessions, the content of which was created taking into account their weaknesses in language assessment. They filled out KWL charts and reflection reports for each session and attended the interviews at the end of the process. As a result of the analysis of the obtained data, nine areas of development emerged. Based on Figure 6, it

can be said that the area in which teachers develop themselves the most is the area related to assessment techniques. Then, this development is followed by cognition developments in test construction and writing test items, assessment results, basic subjects of assessment, assessment principles, grading and scoring, reliability and measurement error. The developments in the cognitions of the teachers after the RRD sessions were detailed in the following areas.

4.4.1. Teachers' Developments Concerning Assessment Techniques

In this theme, as a result of the RRD sessions, teachers' self-perceived developments related to assessment techniques in their cognition were examined in detail. This emerging theme consists of six sub-themes such as teachers' development regarding performance assessment and rubric, teachers' developments in knowledge of alternatives, teachers' developments in specific alternative techniques, teachers' developments in what to consider while developing, choosing or adapting a technique, teachers' developments about assessment technique, and teachers' developments in purpose of using assessment techniques. These sub-themes are explained in the following sections, supported by quotations from relevant data sources.

4.4.1.1. Teachers' Developments Regarding Performance Assessment and Rubric

This sub-theme gives information about participants' development about what performance assessment actually is and its implementation in detail. In addition, their cognition development about what a rubric is, types of rubric, how to create and adapt a rubric according to assessment purposes, and what the rubric provides in assessment applications are also presented in this sub-theme. Table 67 examines each code that makes up this sub-theme.

Table 67.

Teachers' Developments Regarding Performance Assessment and Rubric

Code	f	Code Description
Increased knowledge in performance assessment	38	Development in participants' cognitions about performance assessment
Increased knowledge in rubric design and adaptation	28	How to create a rubric? What are its components? Developments in participants' cognitions considering creating a rubric along with its types.
Development in frequency of using rubric in assessment	25	Developments in participants' cognitions considering using rubric in assessment
Increased knowledge in advantages of using a rubric	23	What does using rubric with performance assessment provide with? What are its advantages? Developments in cognitions considering advantages of using rubric.

Participating instructors attended a RRD session on performance assessment and rubric, filled out a KWL chart and wrote a reflection report. Afterwards, they were asked questions during the interviews. As a result of the analysis of their statements, it was found that there was a development in their cognitions about performance assessment. To illustrate, the participants stated that they had previously only associated the performance assessment with the speaking skill, that they did not have such detailed information about the implementation of the performance assessment, and that they did not even use rubrics in this process. However, after the session, they indicated that they learnt how the performance assessment was applied, how an authentic task could be created for that assessment, the necessity of conducting a pilot study, and the importance of using rubrics in performance assessment. To repeat, based on these statements, it can be said that the participants developed their cognitions about performance assessment. The following excerpt illustrates the situation better:

How would I apply the performance assessment? I would definitely set the objectives first. I would set grading criteria. I would prepare a rubric to grade the performance. In this way, I would prepare the task. For example, I could prepare a communicative task. In this way, I would assess the students who exhibited performance while doing the task in line with the objectives. (Last-post-semi-structured interview, P2).

In addition, the participants stated that they had not used rubric, and when they had been required to use it, they had downloaded it from the internet without considering anything. After the RRD session, they stated that they had learned about the advantages provided by use of rubric in assessment process. Therefore, they would definitely use it when they deem necessary both in performance assessment and other assessment applications so as to obtain more reliable, defensible and evidence-based results. The following excerpt shows the situation:

After using rubric, my results will be reliable and valid. It will allow me to obtain quality data. For example, the results will contain as few measurement errors as possible. Effective wash-back can be enhanced to the student. It can help me in terms of giving feedback. (Comparison pre-semi-structured interview, P3).

As a result, all of the participants stated that they had not used rubrics before the RRD session, but that they would increase the frequency of using rubrics in their assessment applications, as exemplified below:

After the RRD sessions, I learned the importance of using rubrics and from now on, I will definitely use rubric wherever necessary in my assessment practices. I will not ignore using it. (Participant reflection report, P4).

4.4.1.2. Teachers' Developments in Knowledge of Alternatives

As a result of the sessions, it was revealed that the participants showed development in their cognitions about alternative assessment. The participants acknowledged that they learned how alternative assessment emerged, the benefits that it provides, the application of each alternative assessment technique along with its advantages, and that they would often include alternative assessment in their future applications. The following Table 68 examines each code that makes up this sub-theme, as can be seen in detail:

Table 68.

Teachers' Developments in Knowledge of Alternatives

Code	f	Code Description
Development in advantages of alternative assessment	37	Developments in participants' cognitions about alternative assessment: what are the advantages of alternative assessment?
Increased knowledge about alternative assessment technique	34	Developments in participants' cognitions about alternative assessment: development in various kinds of alternative assessment techniques
Development in knowledge of alternatives	31	What is generally alternative assessment? Development in participants' cognitions about general knowledge about alternative assessment.
Integrating alternative assessment techniques in assessment practices	2	Development in integrating alternative assessment in teaching / assessment practices. Putting alternative assessment into practice in classroom while assessing the students.

As a result of the analysis of the statements of the participants, it can be said that they developed their cognition about alternative assessment. When they were asked to talk about alternative assessment before, the participants talked superficially about a few alternative assessment techniques. After RRD sessions, however, participants gave detailed information about alternative assessment. According to the participants, alternative assessment is more advantageous than traditional assessment, as it considers individual differences, allows skills to be assessed and evaluated, and supports students on the way to becoming autonomous learners. The following quote exemplifies this situation:

*I will often include alternative assessment methods because I think that I can more easily manage to consider individual differences and provide construct validity through alternative assessment. I've learned a few advantages of alternative assessment, these are just two of them. In my opinion, making the student an **autonomous learner** and making a **valid assessment** on the basis of skills are its best advantages. (Last post-semi-structured interview, P1).*

Based on the statements of the participants, another self-perceived development is the general knowledge of alternative assessment. Previously, the participants stated that they did not have enough information about the alternatives in assessment and evaluation, they were not competent enough to include them in their practices because they did not know, for instance, enough techniques, they did not know how to plan and apply them. After these sessions, the participant felt competent enough to include alternative assessment in their own

assessment practices. They stated that they have improved themselves in this regard. The following excerpts exemplify this finding:

*S/he knows **alternative assessment** and **how, for what purposes and when to incorporate it into actual assessment practices**. To illustrate, the participant **learned why alternative assessment should be used, its benefits** along with various **alternative techniques**. (Researcher reflection report)*

*I have **learned drawbacks of traditional tests and how alternatives emerged**. I have learned that **all the assessments are not test, but all the tests are assessments**, this sentence makes it clear that there can be something beyond the tests and they are parts of assessment. In other words, alternatives are not tests, but they still assess. Also, I have **learned general characteristics of alternative assessment**, and I can deduce that since they are authentic, they can grab students' attention easily and also they can motivate students. (Participants reflection report, P3).*

As a result, the participating instructors stated that after these RRD sessions, they would integrate alternative assessment into their actual assessment practices more effectively and more frequently, as clarified in the excerpts below:

*From now on I won't use just paper and pencil tests. **In addition to it, I will put alternative assessment into my actual assessment practices more frequently**. (Participant reflection report, P1).*

4.4.1.3. Teachers' Developments in Specific Alternative Techniques

Participants attended the session, which covered main alternative assessment techniques in detail, and then answered the interview questions aimed at revealing whether there were developments regarding knowledge of specific alternative assessment techniques in their cognitions, or not. As a result of the analysis, it was seen that participant instructors made improvement about each of the alternative assessment techniques. Table 69 provides information about alternative assessment method that the participants have developed themselves specifically.

Table 69.

Teachers' Developments in Specific Alternative Techniques

Code	f	Code Description
Increased knowledge in portfolio	32	Development in participants' cognitions considering alternative assessment technique: development in portfolio
Development of self-assessment	20	Development in participants' cognitions considering alternative assessment technique: development in self-assessment
Development in classroom observation	19	Development in participants' cognitions considering alternative assessment technique: development in classroom observation
Development in journal as an alternative technique	7	Development in participants' cognitions considering alternative assessment technique: development in journal
Increased knowledge in peer assessment	4	Development in participants' cognitions considering alternative assessment technique: development in peer assessment.
Conference as an alternative technique	4	Development in participants' cognitions considering alternative assessment technique: development in conference

The first alternative assessment technique that is claimed to be improved by the participants is the portfolio. It was found that the code referring to the development of cognition about the portfolio has the highest frequency rate. Before the RRD sessions, some of the participants mentioned that they only knew the portfolio superficially and applied it in their lessons, while others said that they did not know anything about the portfolio and could not integrate it into their practices. But after the RRD sessions, the participants stated that they had acquired sufficient knowledge about portfolio assessment and could now easily incorporate it into their practices in a professional manner as shown in the excerpts:

Portfolio is not something that can be applied in a lesson, so you couldn't observe me applying the portfolio in the lesson. Portfolio is a method that can be applied continuously throughout a certain process. For example, for a period. I learned that a portfolio is not something to be decided and implemented right away. It is necessary to prepare in advance. What and how will we include in the product file, at what periodic intervals will we check the products, will we grade these portfolios or not? If we are going to give, it is necessary to determine the assessment criteria. After determining these, you need to go into practice. As I said, I will ask the administration for a writing class next year. Then, with a good planning before the lesson, I make a good practice during the semester. (Last post-semi-structured interview, P1).

Another development that seems interesting to the participants is about classroom observation. Instructors stated that they included classroom observation in their assessment practices and that they benefited from it effectively. After the RRD session they attended, the participants confessed that they actually did not know classroom observation and that what they did was full of mistakes. As a result of the analysis, it was revealed that the participants now have enough knowledge about classroom observation. They report to implement it in their actual assessment practices as seen in the excerpts:

I just explained, it is not possible to evaluate the whole class in classroom observation at the same time. If I had been to use the classroom observation assessment technique in the lesson you came for observation, I would have identified certain students, of course, I would have set certain objectives. What I want to see in those students, what behaviour I want to assess, I would have assessed them in the light of rubrics or a check-list and grading criteria. (Last post-semi-structured interview, P2).

I learned that the observation I applied was not actually a real classroom observation. My knowledge has also changed completely in terms of observation. I learned that beforehand, I should determine the behaviors that I will observe, I should determine the students that I will observe, the number of these students should not be too many. I should observe them in class with the help of rubric (Last pre-semi-structured interview, P3).

Last but not least, the participants stated that they heard about conference, an alternative assessment technique, for the first time and learned it thanks to RRD session. Also they reported that since this technique could provide with positive washback to students and teachers through constructive feedback, they would definitely use it in their future applications where they deemed necessary. The following excerpt exemplifies this view of the participants:

Conferences and interview can be used. This technique suits well while teaching writing. In conferences, teachers come into direct conversation with students. Conferences provide one to one interaction. Teachers can give direct feedback. Actually, it is not practical for teachers. It takes time. It can be tiring for teachers. But when I think of it, it is really helpful. It increases positive washback. It has really advantages. It improves. It is formative. (KWL chart, P3).

4.4.1.4. Teachers' Developments in What to Consider while Developing, Choosing or Adapting a Technique

Participants' self-perceptions about what they should pay attention to when choosing or developing an appropriate assessment technique or assessment tool for their own assessment purposes, teaching context and student profile are presented in detail in Table 70 below. Before the sessions, the participants stated that they only considered practicality and content validity when choosing any technique or tool for their practice, but after the sessions they attended, they were enlightened about this issue and from now on they would make more informed choices in their applications. Detailed explanation for each code belonging to this sub-theme is presented as seen in Table 70:

Table 70.

Teachers' Developments in What to Consider while Developing, Choosing or Adapting a Technique

Code	f	Code Description
Considering washback	15	Development in washback while creating or selecting an assessment tool. It should provide with positive washback both for teachers and students.
Considering validity	14	Development in validity while creating or selecting an assessment tool / technique. It should be valid.
Considering reliability	14	Development in reliability while selecting an assessment technique. It should be reliable
Considering authenticity	9	Development in authenticity while selecting an assessment tool / technique. It should be authentic
Considering student profile	7	Development in participants' cognitions considering students' profile while creating or selecting an assessment tool / technique.
Considering objectives	4	Development in considering objectives while creating or selecting an assessment tool.
Testing a test	3	Development how to test appropriateness of a test.
Considering teacher's own knowledge	3	Development in participants' cognitions considering teacher's own knowledge while creating or selecting an assessment tool. Teachers should know how to use administer it to the students.
Considering measurement error	2	Development in considering measurement error while creating or selecting an assessment tool. It should be error free.

First of all, it was stated by the participants that they made progress about washback while choosing or developing any assessment tool or technique. The participants stated that they had no knowledge of what positive washback was before, but after the session they

attended, washback became one of the main factors to be considered in choosing a technique or assessment tool for them. According to the participants, the assessment technique or tool used should provide washback to both the student and the teacher. Thus, students' learning and teachers' teaching can be brought to a better level. The following excerpt exemplifies the cognition development of the participants:

*I didn't know washback. I didn't have any idea about washback. I read and understood that any assessment technique and assessment method or any exam, **test means nothing without washback**. As long as it enhances positive washback to students, then **it works well and it facilitates and improves learning**. That's why the very first thing that I consider is washback (Participant reflection report, P4).*

Later, the participants expressed that reliability should be taken into account when choosing an assessment technique. The participants, who mentioned only practicality before the RRD sessions, now state that reliability had a very important place in technique selection or assessment tool development. Participants stated that the assessment tool or technique used or developed must be reliable and enable reliable results to be obtained. In fact, participants who understand how important reliability is in order to ensure validity, therefore, definitely consider reliability in choosing a method. The participant statement quoted below exemplifies this finding as seen:

*I consider reliability when choosing the appropriate assessment tool or technique for my purposes. **Will I be able to assess in a reliable way? Will I be able to accurately identify the existing potential in the student? We have to consider many things**. But I think the most important thing is reliability. Because, if reliability is not provided, validity does not occur. Thus, what has been done becomes in vain. For this reason, I try to create a reliable test or choose a reliable method or assessment tool to make an accurate assessment and evaluation. (Comparison pre-semi-structured interview, P1).*

In addition to these, it was stated by the participants that knowledge and implementation ability of the teacher was one of the issues to be considered in choosing an assessment tool or technique suitable for their purposes. According to the participants, a teacher who wants to include any assessment tool in his/her real practices should have sufficient knowledge about implementation of that technique and be skilled to use it appropriately. The participants advocate the view that teachers should have enough knowledge about all of the various assessment methods and that they should apply whichever is appropriate for them in the lesson. Excerpt taken from the interview clarifies this finding.

*The assessment technique I will choose or the assessment tool I will develop **should appeal to me. Will I be able to apply it? Do I have enough knowledge and skill level to apply it correctly?** I am also looking for answers to these questions. (Comparison pre-semi-structured interview, P4).*

4.4.1.5. Teachers' Developments about Assessment Technique

This sub-theme shows the developments in the cognitions of the teachers in terms of assessment techniques in general. These self-perceived developments in participants' cognitions are related to participants' ideas about shortcomings of classical assessment techniques, their self-perceived criticisms of assessment techniques used by them before the RRD sessions. Also, information about an increase in the number of known and applied assessment techniques is provided. Table 71 presents codes belonging to this sub-theme in detail, along with operational definitions and extracts from data sources.

Table 71.

Teachers' Developments about Assessment Technique

Code	f	Code Description
Development in the knowledge of kinds of assessment techniques	18	Development in participants cognitions considering general knowledge about assessment technique, kinds of assessment techniques
Increased knowledge about evaluation of techniques being used	10	Development in assessment techniques, development in knowledge of pros and cons of assessment technique that they are currently using, being able to evaluate effectiveness of currently used assessment technique.
Development in implementation of assessment technique in practice	5	Development of implementation, practice, of any assessment technique in actual assessment practices.
Drawbacks of traditional assessment	4	Development in traditional assessment techniques. Being able to evaluate effectiveness of current traditional assessment technique.

As a result of the analysis, it has been revealed that the knowledge of the participants about the assessment technique has increased in general, and accordingly, they see themselves as more competent in determining and using a valid and reliable assessment method suitable for their objectives. This emerging result is elaborated through the excerpts:

*You could not observe me using the assessment technique in your previous observations and in your observations after these sessions, **but there is a great difference between my knowledge level before the sessions and my knowledge level after the sessions. I knew the multiple choice test as an assessment technique. But this is generally a sit-down written exam. Multiple choice is a kind of test item. There***

*has been an **improvement in my knowledge about assessment technique**. Traditional techniques, contemporary techniques... I learned what they are and how they are applied. I learned what they are used for, etc. (Last post-semi-structured interview, P1).*

*As a result of these sessions, I will use alternative techniques to improve the student's learning in the process. I will not conflict with our institution, I will do what it says, but by **using these different and alternative assessment techniques**, I will both support the students' learning and try to improve my own teaching. (Last pre-semi-structured interview, P2).*

In addition, another finding showing that participants' developments in their cognitions about level of knowledge related to assessment techniques is that the participants evaluated their previous assessment techniques by indicating their strengths and weaknesses. In other words, the participants stated that their level of knowledge about assessment techniques increased, that they know which technique had advantages and disadvantages, so they could criticize themselves and the techniques they had used before the RRD sessions. According to them, the techniques used were insufficient to reveal individual differences, weak in terms of construct validity, and made students passive in the process. The following quote exemplifies this situation as it can be seen:

*Prior to these sessions, I generally used written sit-down exams, one of the classical assessment techniques. Its strength is its practicality. I can criticize their weak points after these sessions. First of all, **these tests ignore individual differences among students**. We are language teachers; we also assess students' language skills. **While assessing these skills, how can we do this with classical assessment methods? This is very difficult or even impossible**. Therefore, I better understand the title of the unit I read about the alternative assessment method. "Beyond the tests: alternatives in language assessment." In this way, **alternatives reach what classical tests cannot reach**. (Comparison pre-semi-structured interview, P3)*

4.4.1.6. Teachers' Developments in Purpose of Using Assessment Techniques

With this sub-theme, the aims of the participants to use assessment techniques are stated. The participants, who used techniques only to determine learning status of the students before the RRD sessions, stated that, after the RRD sessions, they have started to use assessment techniques for school improvement, to evaluate their own teaching or to give

washback to the students and motivate them. Table 72 clarifies by detailing for what purpose the participants use the assessment techniques:

Table 72.

Teachers' Developments in Purpose of Using Assessment Techniques

Code	f	Code Description
Using assessment techniques for evaluating teaching	5	Development in cognitions considering purposes of using any assessment technique: Using for evaluating effectiveness of teaching.
Using assessment techniques for positive washback	4	Use of assessment techniques for providing positive washback to students by giving feedback also to get positive washback as a teacher
Using techniques for identifying weaknesses and strengths of students	3	Development in cognitions considering purposes of using any assessment technique: Using assessment technique for identifying students strengths and weaknesses
Using assessment technique for school improvement	1	Using assessment techniques to lead school to improve itself or to take a step to improve quality of school.
Using assessment techniques for motivating students	1	Participants' self-perceived development: using any kind of assessment technique in order to motivate students.

When Table 72 is examined, it is clearly seen that the participants have improved themselves concerning purposes of assessment technique usage. Before the RRD sessions, participants used assessment techniques only to determine whether students have learned the subject taught to them or not. However, participants reported to use assessment techniques for various purposes after the RRD sessions. One of them is to use the techniques to reveal the strengths and weaknesses of the students. For this purpose, the participants stated that they would identify weaknesses of the students and provide them with washback, and then they would motivate them by revealing strengths of the students.

After these sessions, I can use assessment and evaluation techniques to determine the strengths and weaknesses of the students. (Last pre-semi-structured interview, P3).

Portfolio, interview, peer assessment, observation, self-assessment, journal, etc. We give feedback to the students about their weaknesses in the learning process by considering student differences in these assessment practices. This increases a "washback" effect. This is a very useful thing. For this purpose, I deliberately use the assessment techniques. (Last pre-semi-structured interview, P4).

4.4.2. Teachers' Developments in Writing Test Items and Test Construction

As a result of the content analysis, detailed information about the development of participant instructors' cognitions about test preparation and test item writing is given. Participants' self-perceived cognition development about test construction and item writing were categorized under five headings such as teachers' developments in writing kinds of test items, teachers' developments in content of a test in test preparation, teachers' developments in what to consider while writing test items, teachers' additional developments regarding test preparation, teachers' developments considering assessment principles in test preparation. Each of them has been explained separately in the following sections.

4.4.2.1. Teachers' Developments in Writing Kinds of Test Items

As a result of the analysis of the answers given by the participants to the interview questions, expressions they stated in the KWL chart and reflection reports, development in their cognitions related to writing test item is presented under this sub-theme. While preparing the test item before the RRD sessions, the participants stated that they either wrote the test item or downloaded it from the internet without considering anything, regardless of the type of item. In addition, the participants, who stated that they did not know writing rules of any specific test item type, were found, as a result of the analysis, to have improved themselves in both test item writing in general and the specific test item type writing. Detailed information about this sub-theme is given in Table 73:

Table 73.

Teachers' Developments in Writing Kinds of Test Items

Code	f	Code Description
Increased knowledge in kinds of test item	42	Development in participants' cognitions considering general kinds of test items and how to write them.
Writing multiple choice test item type	32	Development in writing test items, multiple choice test item construction, considerations for writing a multiple choice test item
Development in writing t f test item	8	Development in participants' cognitions considering writing test item, true false item type
Development in writing essay type test item	4	Development in writing kinds of test item type, development in writing essay type test item.
Development in writing matching test item	2	Development in writing test item types, development in writing matching test item.
Writing short answer test item	2	Development in writing kinds of test items, writing short answer test item.
Development in writing fill in the blanks item type	1	Development in writing kinds of test items, writing short answer test item.

It is seen from Table 73 that the subject in which the participants showed the most improvement is writing the multiple-choice test item. Before the RRD sessions, the

participants stated that they only paid attention to the content validity and item difficulty level while preparing a multiple-choice test item. After the RRD sessions, the participants stated that they learned a lot about writing multiple-choice questions. The participant instructors claimed that the most striking feature of multiple-choice test items was that alternatives and distractors have a diagnosis mission, and reported improving themselves in writing multiple-choice test items after these sessions. These explanations below explicate the finding better:

*The article also has provided that in a question of multiple choices, **all alternatives should be about the same theme**, the teacher shouldn't prepare a test item's alternatives form different themes or unite. I have found out that, I have never heard it before, **a true/false test item shouldn't include some certain words such as "always, never, sometimes, usually, because these kind of test items lead the students to guess the correct answer.** Such test items don't serve to objectives; all they do is to trigger the students to pick up a strategy for test items. (KWL chart, P2).*

*We discussed how to write a multiple choice test. I learned that this job is really hard. **I learned new information both in terms of the view of the question and in terms of the functions of the alternatives in the question. For example, we should pay attention to wording.** We should avoid precise statements and **not use words such as always, never. The options must be the same length.** I did not know this, for example, if there are 5 options in a question, all of these options must be from the same unit or even the same theme. Here, **the answers should not follow a certain order so that the student does not guess.** Because this destroys the reliability. The most important thing I learned is that the options we call distractors also have a diagnostic function. I learned these. I can prepare a multiple choice test. (Comparison pre-semi-structured interview, P1).*

In addition, participants stated that they had limited knowledge in terms of type of test item that they added to their exams. However, after the RRD session they attended, the participants stated that they acquired enough information about various test items and that they would include these new types in test construction process from now on. The following excerpt illustrates the situation better:

*There are subjective tests and objective tests. There are test items that belong to them. **I know what these types of items are. There are rules to be followed** while writing them, **I learned them.** I can apply (Comparison pre-semi-structured interview, P4).*

4.4.2.2. Teachers' Developments in Content of a Test in Test Preparation

In this sub-theme, detailed information about the teachers' self-perceived developments regarding the content of the test and its design during test construction process is presented. To be clear, this sub-theme deals with teachers' self-perceived development after RRD session concerning what they should consider when choosing a test item for a test, how the order of the test items they decide to use should be, what kind of test items they should include in a test according to their purpose, and how a test should be designed. Table 74 details this sub-theme:

Table 74.

Teachers Developments in Content of a Test in Test Preparation

Code	f	Code Description
Development in considerations in selecting test items for a test	24	Development in participants' cognitions about what to consider selecting proper test items in a test
Considering sequence of items in preparing tests	14	Development in participants' cognitions considering the sequence of test items while including them in a single test.
Development in deciding on kind of test item to be included in	10	Development considering necessary kinds of test items which should be included in a single test.
Considering design of a test in test preparation	10	Development considering organizing design of a test.
Including various kinds of test items in a test preparation	6	Development about including multiple kinds of test items in a single test.

When the Table 74 is examined, the most frequently expressed self-perceived development of the participants is considerations while selecting test items for a test. The participants, who had previously considered only practicality and content validity, stated that they would consider washback, measurement error, reliability, authenticity and, most importantly, students' profiles when choosing appropriate test item for a test after the RRD sessions. Based on these statements, it can be said that the participants showed development in selecting appropriate test items for a test during the test construction process as it is demonstrated in the following excerpt.

While I am selecting a test item for a test, I can evaluate item candidates by asking some questions. I've learned some principles which can be used while evaluating the tests. These principles are practicability, reliability, validity, authenticity and washback. I will of course take into consideration while choosing a test item for a test. (KWL chart, P3).

Next, the participants learnt that during the test construction process, they first should determine type of test they are aiming to prepare, and accordingly, they should select test items suitable for the test type and include them in the test. In addition to these, as a result of

the analysis of teachers' own words, it was revealed that they also showed development in deciding on the type of item for a test. While they included the question type in a test randomly beforehand, after the RRD sessions, the participants stated that they primarily considered both the test type and the assessment objectives when deciding on the question type for a test. According to them, these two factors help the assessors to decide on the test item type. Following extract illuminates the situation:

If I am going to assess the student in the recall of the information dimension, I choose short-answer questions. If I want them to use higher order thinking skills, I use the open ended question type. So when choosing the question type, I first consider my objective. Depending on my objectives, the test item type that I will prefer becomes clear by itself. (Last post-semi-structured interview, P1).

These results suggest that after the RRD sessions, teachers became more conscious and more systematic while preparing tests.

4.4.2.3. Teachers' Developments in What to Consider while Writing Test Items

This sub-theme is related to the development of the participants in writing test items. To be more specific, it provides information about what participants will consider when writing a test item for the test they will use in their assessment practices. In other words, after the RRD sessions, the teachers learned to write questions according to the table of specifications and Bloom's taxonomy while writing the test item. Teachers who are aware that they can increase validity in their assessment practices developed their cognitions also in using wh-questions correctly while writing questions, choosing appropriate vocabulary during writing test items, and writing questions that will measure higher-order thinking skills respectively. Table 75 details this sub-theme:

Table 75.

Teachers' Developments in What to Consider while Writing Test Items

Code	f	Code Description
Incorporation of Bloom's taxonomy in test item writing	18	Development in taking Bloom's taxonomy into consideration while writing any kind of test item.
Using suitable wh-question while writing test items	10	Development in being careful while using wh-questions while writing any kind of test item.
Considering vocabulary while writing test items.	7	Vocabulary usage. Development in using words consciously while writing test items. Development in being aware of power of words in item writing.
Increased knowledge in enhancing validity in item writing	6	Development in how to ensure validity in item writing.
Development in item writing measuring higher-order thinking skill	5	Development in participants' cognitions in writing test items making students use their higher order thinking skills.
Importance of wording in item writing on students' learning	4	Development in being aware of influence of wording and vocabulary used in test items on students learning.
Taking table of specifications in writing test items	1	Development in knowledge of table of specifications and employing it while writing test items.

First of all, it was previously found out that although some of the participants knew Bloom's taxonomy, they did not use it while writing test items. Furthermore, some of the teachers were found to know only name of Bloom's taxonomy, and naturally not to use it in writing test item process. However, after the RRD sessions, it was revealed that the participants learnt about Bloom's cognitive taxonomy and had sufficient knowledge on how to integrate it into test item writing process. Also they claimed to prepare test items assessing students' higher order thinking skills. The direct quotation from the data below illustrates this situation:

We learned the levels in Bloom's taxonomy. I learnt its last revision. We will take its steps into account when writing questions. For example, I realized that most of the test items I prepared require answers in the dimension of knowledge. But I also learned that I can make students use their higher reasoning skills with the questions I prepare. For example, I can provide analysis with open ended question type. I will consider bloom's taxonomy to write a quality test item. (Comparison pre-semi-structured interview, P1).

In addition, the participants stated that they did not know that much attention should be paid to the words used during writing test item. They reported learning thanks to the RRD session that wh-question or other words used in the question were effective on the students' learning performances, and that they would be more careful with the words used in test items

to direct students to meaningful learning rather than memorisation while writing the test items. The following quote illustrates the situation:

The article that I read also enabled me to learn that we should not use “what, who, when” so frequently in our test items because they directly ask for knowledge without letting the students use their analytic, interpretation skills. I have learnt that such questions may lead the students to rote learning, namely, the students after a while memorize things, and it cannot be meaningful learning. Therefore, I should ask more elaborated question when preparing test items so that the students are able to present and use their higher thinking skills. (KWL charts, P2).

Last but not least, the participants claiming to prepare test items which require reasoning at knowledge level learnt that test items in a test should enable students to use their higher order thinking skills. After the RRD session they attended, the participants explained that while they were constructing a test and writing test items, they would include questions which require using higher-level thinking skills. This finding is explicated in the following excerpt:

Previously, when I was preparing a test or writing a test item, I used to prepare questions that could assess things like memorization, not the student's higher level knowledge. But I learned that questions that require high-level thinking skills should also be asked. Analysis, synthesis... If I think about it, I was creating questions that assessed students' lower level skills. But I realized later, thanks to these activities and readings that I need to use questions that the student constructs knowledge while answering them. I will be more careful from now on. (Last pre-semi-structured interview, P4).

4.4.2.4. Teachers' Additional Developments Regarding Test Preparation

With this sub-theme, participants' cognition development in test construction and item writing as a result of the RRD sessions is given. This development is related to ethical issues in test preparation and test item writing, measurement error in test preparation, ideal time to prepare a test, and avoiding ambiguity while preparing test and writing items. Table 76 shows the detailed information of the codes belonging to this sub-theme:

Table 76.

Teachers' Additional Developments Regarding Test Preparation

Code	f	Code Description
Increased knowledge in test types	27	Development in kinds of test formats
Preventing ambiguity while preparing a test	11	Development preventing ambiguity for understandable instructions
Considering ethics in test preparation	6	Ethical issues in test preparations. Development in how to ensure ethics while preparing any kind of test format.
Ideal time for test construction	3	When to prepare a test? Development about when to construct a test for exams.
Considering measurement error while preparing tests	2	Measurement error. Development considering measurement and its influence while preparing tests.

When the table is examined, it is clearly seen that there is a development in cognitions of the participants about the test formats. Participants, who did not consider the test format when choosing or constructing a test for their own assessment practices, expressed that they had acquired information about the topic during the RRD sessions and that they would consider test format when deciding on the test item to use. The excerpt, which exemplifies the development of the participants about the test format, is given below:

*I have learnt about **subjective and objective test items** that subjective tests include short-answer and essay type test items are subjective tests and true/false, matching and multiple choice tests are types of objective tests. (KWL chart, P3).*

From the statements given by the participants, we can suggest that teachers have improved themselves in considering ethical issues in test preparation process. When ethical issues were mentioned before the RRD sessions, the participants used to associate being fair and objective in grading with ethics. Apart from it, they did not know anything related to ethics. As a result of participating to RRD sessions, participants reported to learn how to consider ethics in test preparation. For example, they stated that it is an unethical to ask a question whose subject is out of content. In short, EFL instructors were found to have developed their cognitions considering the issue. The following extract exemplifies the situation better:

As a result of the sessions, I learned new things about how to consider ethical issues while preparing exams. Behaviours such as asking very difficult questions to punish students or asking easy questions to reward students are unethical behaviours and they should be avoided in test construction process. I can ensure ethics by behaving in this way while preparing a test and writing a test item. (Comparison pre-semi-structured interview, P1).

As a result of the analysis, it was found that the participants developed their cognitions about the ideal time to prepare a test. Previously, instructors who prepared the exam and administered it to the students a day before or a few hours before the exam stated that exam preparation time should be long enough before the administration of the exam. According to them, the exam they prepared should be checked again and its suitability for practice should be tested. For this, it is necessary to prepare the exam long enough before applying it and then it can be administered to the students. The following extract exemplifies the situation better:

As soon as I prepared the test, I would not administer it immediately. A few days after preparing the test, I would revise it again and even ask a colleague to check it out. Then I would run the test. (Last post semi-structured interview, P3).

4.4.2.5. Teachers' Developments Considering Assessment Principles in Test Preparation

Participants learnt assessment principles such as reliability, validity, practicality and washback as a result of the RRD sessions they attended. The cognition developments of the participants regarding how these principles are integrated into test preparation process are presented in this sub-theme. Table 77 presents detailed information about the emerging results:

Table 77.

Teachers' Developments Considering Assessment Principles in Test Preparation

Code	f	Code Description
Considering validity types in test preparation	17	Development in participants' cognitions about incorporation of assessment principles in test construction: validity
Reliability in test preparation	17	Development in participants' cognitions about incorporation of assessment principles in test construction: reliability
Considering washback in test preparation	3	Development in participants' cognitions about incorporation of assessment principles in test construction: washback

The majority of the participants' cognition development related to considering assessment principles in the test preparation process is about washback. According to them, tests that will be administered to the students should increase washback both to the students and to the teachers, and it will be beneficial in the decisions taken to create better learning and teaching opportunities after the application of the tests. For this reason, the teachers indicate that they will consider the necessity of a test to provide washback to stakeholders when preparing the tests. The sample taken from the interview illustrates the situation:

*The test I prepared should **have a positive effect on the student**. After applying the exam, we will announce the results to the students. The test we prepared, while announcing **the exam results**, should be qualified in terms of giving feedback to*

students about the their strengths and weaknesses. The test I will prepare or adapt should be an exam with appropriate content in terms of washback (Last pre-semi-structured interview, P1).

In addition, the results of the analysis revealed that the participants also developed their cognition on how to ensure reliability while preparing tests. Before the RRD sessions, the participants stated that they only prevented students from cheating in order to ensure reliability. However, after the RRD sessions, the participants learned that the test which will be administered to the students also affects the reliability. Moreover, they stated clearly that while preparing a test for this purpose, they would pay attention to giving clear directions to students, writing style along with size of letters, not to cause ambiguity and not to give a clue about the answers to the questions in the exam. This finding is better explicated in the following:

*To be more precise, concerning about validity or reliability for the test preparation was not a goal of mine. More or less, **I may pay attention to validity while preparing the test but to be honest, I didn't focus on reliability.** I learned that we need to pay attention to reliability while preparing tests. It affects test **validity**. I know how it will affect it, the most important thing is **to use a clear and understandable language**. From now on, reliability will be ensured while preparing the test. (Last-pre-semi-structured interview, P3).*

4.4.3. Teachers' Developments Regarding Assessment Results

With this theme, the results that refer to the cognition development of participants in the issues related to the assessment and evaluation results are presented. The theme consists of five sub-themes which are roughly about use of assessment results, sharing assessment results, interpreting assessment results, decision making as a result of assessment, reflection of assessment results. Each sub-theme is examined in detail in the following sections.

4.4.3.1. Teachers' Developments in Use of Assessment Result

This sub-theme gives information about purposes and areas of use of the results obtained by the participants through assessment practices. The participants used assessment results before the RRD sessions almost only to determine whether the students understood the subjects or not and then to decide on their success. After the reading and discussion sessions, EFL instructors stated that they would use the assessment results in different fields which are detailed in Table 78 below:

Table 78.

Teachers' Developments in Use of Assessment Result

Code	f	Code Description
Using assessment results for students' development	24	Development considering using assessment results for the development of students.
Using assessment results for teacher development	22	Development considering using assessment results for developments of teachers.
Using assessment results for adapting assessment tool	14	Development in cognitions considering using assessment results for development of assessment tool or technique which was employed.
Using assessment results for school improvement	13	Development in cognitions considering using assessment results for better school

By examining results, it can be seen that cognition development of the participants regarding the use of the assessment results is mostly related to students. Participants stated that after the RRD session, the instructors would use the results for purposes such as motivating students, identifying their strengths and weaknesses, and providing them a washback through constructive feedbacks. Therefore, they believe that they can help students improve their learning. The following excerpts illustrate the situation:

I would identify the weak and strong parts of the student. Evaluation results are somewhat diagnostic. I used to decide what the student could and could not do. In line with this, I would aim to improve the students. (Last post-semi-structured interview, P2).

Considering assessment results, I will take students into consideration immediately. I will discover my students' strengths and weaknesses. I will praise and motivate my students considering their strengths, and I will discover their problematic areas and I will treat them. I have learned that assessment results are important and I can enhance with washback to my students through them. (Participant reflection report, P3).

One of the purposes of participants' use of assessment results is to evaluate teachers' own teaching. Almost all of the participants stated that they used the assessment results to check students' understanding levels. However, after the RRD sessions, the participating instructors added that they would use the results to reveal their professional development needs, that is, to evaluate how effective they were in teaching, and in which part(s) of

teaching they need further support. The following quote taken directly from some of the participant's explanations show the situation:

The grades besides informing the students' developments, they enable the teachers to improve and design their own further teaching practices and to set down next steps for better teaching procedures. (Participant reflection report, P2).

*When I look at the assessment results, if the **majority of the class got low grades, the problem may be stemmed from my teaching.** That's why I would teach by reducing my lessons to a little more basic level. I would change **my own instruction according to the assessment results and adapt the materials, activities and even assessment tool according to the results.** (Post-semi-structured interview, P4)*

In addition, the participants stated that they could change or modify the assessment tool by using assessment results. The participants, who did not mention any use of results focusing on adapting tool stated that after the RRD sessions, they would use assessment results to evaluate the assessment technique or tool that they utilized in their practices and make changes in it when it is necessary. Therefore, they believe that they could increase the quality of obtained results and naturally they can take quality decisions accordingly. The following excerpt explicates this finding:

*By using the assessment results, I can **change the assessment technique or the assessment tool I use.** I did not know that the assessment result could be used for this purpose. **Maybe the tool is not valid.** The student has that competency, but that assessment tool does not help the student to reveal it. In this context, **I edit the tool or change it completely.** (Last post-semi-structured interview, P1).*

4.4.3.2. Teachers' Developments in Sharing Assessment Results

This sub-theme presents findings referring to EFL instructors' cognition developments of participants regarding announcing assessment results. It was revealed that after attending the RRD sessions, participants generally learnt how to share assessment results with stakeholders. Their development was found to be related to generally announcing the assessment results, being careful in the selection of words while sharing the results, and considerations which should be kept in mind while announcing the assessment results. Detailed information about the sub-theme is given in Table 79:

Table 79.

Teachers' Developments in Sharing Assessment Results

Code	f	Code Description
Development in announcing assessment results	29	Development in informing students and other stakeholders about assessment results.
Caring about wording while sharing assessment results	6	Development in participants' cognitions about use of vocabulary – wording- while sharing assessment results.
Principles in sharing assessment results	5	Development in participants' cognitions about principles can be taken into consideration while sharing assessment results.
Considerations while informing stakeholders about results	2	Development in general things which should not be ignored while sharing assessment results.

Based on the results, it can be concluded that the knowledge level of the participants about informing stakeholders as a result of the assessment has increased. This progress is related to methods of sharing assessment, giving students constructive feedback indicating their strengths and weaknesses while explaining the results, in other words, providing washback to the students during informing them about assessment results.

First of all, participants stated that they learned various methods of explaining assessment results and how to make an effective explanation to students through these methods which will have an effective influence on students' learning. The following excerpt explains the situation:

Before this reading and discussion session, I didn't know what exactly communicating results suggests. For instance, I have never arranged a meeting with students to discuss their grades with a detailed report to indicate their strengths and weaknesses. All I did was to declare their grades, praise ones who did quite well and nothing more, no further feedback was offered to clarify why they had that grade. However, I have learnt various ways of sharing assessment results and the fact that the teacher should be honest enough to frame the students' progress namely showing them their strengths along with their weaknesses. Describing their strengths and weaknesses with honesty will help the students to make an improvement. Namely, forthrightness should be owned. (Participant reflection report, P2).

In addition, participants stated that they learned information that they had never known or even heard of, thanks to the RRD session. To illustrate, they did not pay attention to wording while explaining the assessment results to the students; they only informed them politely without offending the students. However, as a result of the RRD session, they learned that wording used while declaring the results or giving feedback depending on the results conveys hidden messages and affects the students' self-perceptions in a positive or negative way. After that, the participants reported that they would be more careful in use of words while explaining the results. The following excerpt exemplifies the situation:

*I have learned that **wording has an important role while sharing assessment results.** Firstly, I have learned that I **should not use sharp expressions** like never, always while sharing assessment results. Also, if it is necessary use modal verbs, **we should also be careful because some modals can give the message to the students that you are incapable of doing anything.** In order not to lead students to feel inefficacy, we **should be careful.** I did not know but I have learned them thanks to this reading and discussion session. (Participant reflection report, P4).*

4.4.3.3. Teachers' Developments in Interpreting Assessment Results

This sub-theme includes findings of the study that refer to participants' cognition developments related to the interpretation of assessment results. This development is about item analysis as a statistical process that helps to make interpretation effectively, and in which areas and for what purposes the interpretation of assessment results should be made. Detailed information about the sub-theme is given in Table 80:

Table 80.

Teachers' Developments in Interpreting Assessment Results

Code	f	Code Description
Increased knowledge in item analysis	19	Self-perceived development in what item analysis is
Interpreting assessment result focusing on students	7	Development in cognitions about interpreting data to provide students with washback
Interpreting assessment results for teacher efficacy	6	Development interpreting results for evaluating efficacy of teaching.
Interpreting assessment results for school development	3	Development in cognitions regarding interpretation of assessment results for school development.
Interpreting assessment results for the efficacy of assessment tool	3	Self-perceived development in participants' cognitions regarding the interpretation of assessment results for evaluating the efficacy of assessment tool.

According to the results, a large part of the cognition development of the participants is related to a statistical process that helps the teachers in the interpretation of results. Before RRD sessions, participants stated that they did not know this statistical process and therefore they did not use it in their assessment practices. After the RRD sessions, the participants learned about item difficulty level and item discrimination index, and their contribution to interpreting assessment results. Emphasizing the importance of interpreting the results in the interviews, the participants stated that they would definitely consider item analysis, that is, item difficulty level and item discrimination index, when interpreting the assessment results in their future assessments. The following excerpt explicates this finding:

*Item analysis is a statistical process used to **determine the difficulty level of questions in the exam**. The exam should be able to **distinguish between those who know and those who do not**. The exam should **be neither too hard nor too easy**. Thanks to these methods, I can see which question was answered correctly by how many people and how many were answered incorrectly. Thus, I can determine **the subject that the students cannot understand and create a treatment lesson after the exam**. These statistical calculations are very useful, I learned them and I will apply them. (Comparison pre-semi-structured interview, P3).*

A result that is in line with the aforementioned result has emerged as a result of the analysis. The participants stated that by using item analysis, they could identify the strengths of the students, thus motivating them. In addition, the participant instructors stated that they

could determine which question in the exam was answered wrongly by the majority of the students thanks to the item analysis. Also, they claimed that they could identify the subjects that students lacked knowledge about and help them to compensate for these weaknesses in a more beneficial way. The following excerpt supports this finding:

I learned that by interpreting the assessment results, I can find out which subjects the students are knowledgeable about and which subjects they lack of knowledge. Then, in order to bring the students to a better level, I plan a way with them to compensate these weaknesses. (Last pre-semi-structured interview, P3)

Last but not least, the participants stated that until now they hadn't not interpreted the assessment results to improve school, but after the RRD sessions they learnt that assessment results should be used to bring the school to a better position. After the sessions, participants believe that they will determine the deficiencies in the school or the areas that need to be improved, and by correcting them, they will bring the school to a good level and thus increase the quality of education. The following excerpt explicates the situation:

*Looking at the assessment results, I interpret as follows: **Did we, as an institution, meet our expectations?** Suppose that we couldn't... What kind of arrangements can we make within the institution by looking at the results by determining the deficiencies? **What can we do for the better? What should administrators do, what should teachers do, what should students do?** I can determine these, I think, as a result of interpretation, but I need to do this interpretation with all stakeholders. (Comparison pre-semi-structured interview, P4).*

4.4.3.4. Teachers' Developments in Making Decisions Based on Assessment Results

As a result of the analyses, another development in the cognitions of the participants is related to making decisions by looking at assessment results. Before the RRD sessions, the participants mostly made a decision about a student's passing or failing the class or course based on assessment results. However, after the RRD sessions, the participants improved themselves on issues such as what is needed to make a decision by looking at assessment results, and what needs to be done during the decision-making phase. Table 81 presents findings related to the sub-theme:

Table 81.

Teachers' Developments in Making Decisions Based on Assessment Results

Code	f	Code Description
Obtaining quality data for decision making	7	Development in participants' cognition about making decisions based on assessment results which requires first obtaining quality data.
Development in making decisions based on assessment result data	7	Self-perceived development in the issue of general making decisions in assessment
Potential influence of valid decisions on teaching	2	What does valid decision mean? Self-perceived development in participants' cognitions about valid decision based on assessment data.
Making evidence-based decisions based on assessment data	2	Self-perceived development in importance of making evidence based decisions based on assessment data.
Including stakeholders in decision making process based on results	2	How should decisions be made? Self-perceived development in participant's cognitions regarding including all of the stakeholders into decision making process.

The participants were asked questions about decision-making process, taking into account assessment result. After the RRD session they attended, the participants first mentioned the first prerequisite of decision making is collecting quality data. In other words, they reported that teachers had to obtain quality data through assessment practices then they could have made appropriate decisions. Also participants mentioned that they had developed themselves about issue of validity. According to them, they can integrate validity in decision making process and only valid results can lead to valid decisions improving education and training. Invalid assessment results are useless in the decisions made. Decisions taken with these in mind do not improve either teaching of teachers or learning of students. The excerpts that exemplify these two results, which are in the same direction explain this situation:

*Furthermore, data which we use in order to make decisions **should be high in quality**. That is, **we can't say that all the data are useful for decision making**. We can't use whatever data we want. **Quality data are needed**. (Participant reflection report, P3).*

*As a teacher, I learned the importance of obtaining valid results. That is, only valid **decisions made with valid results improve us**. I am aware of that. For this, I need to **obtain valid and error-free assessment results so that I can obtain valid decisions**. As a result, I can improve my teaching and my students' learning through beneficial initiatives. (Participant reflection report, P1).*

In addition, the participants stated that decisions made by considering assessment results are not only about teacher, but students, administrators and even parents should also participate in this process. Previously, only teachers had leading role in decision-making but after RRD sessions they added that opinions of all stakeholders would be included in the decisions taken to increase the quality of teaching. The following excerpt illustrates the situation:

As a result of the reading and discussion session, I have realized that making decision is not an easy task. It poses great responsibility to teacher. Actually, it is not only about teachers. It matters all the stakes holders because decisions made based on assessment result will influence all of them. Also, decisions should be made with a great care. That's why, students teachers, administrators even parents should be included in this process. (KWL charts, P4).

4.4.3.5. Teachers' Developments in Reflection of Assessment Result

Under this sub-theme, codes representing participants' development in their cognition regarding reflection of assessment results is gathered. Table 82 shows detailed investigation of this sub-theme:

Table 82.

Teachers' Developments in Reflection of Assessment Result

Code	f	Code Description
Reflection of assessment results by teachers.	6	Self-perceived development in participants' cognitions regarding reflection of assessment results: reflection from teachers, how should teachers reflect on assessment results?
Reflection of assessment results by students.	5	Self-perceived development in participants' cognitions regarding reflection of assessment results: reflection from students, how should students reflect on assessment results?
Reflection of assessment results by school administrators.	4	Self-perceived development in participants' cognitions regarding reflection of assessment results: reflection from school administrators, how should school administrators reflect on assessment results.

It is seen that results indicating development regarding reflection of assessment results have been classified into three codes such as reflection of teachers, reflection of students and reflection of administrators.

Participating instructors stated that they did not think reflectively about assessment results before. However, after the RRD sessions, participating teachers stated that reflection on assessment results should definitely be done. According to the participants, teachers should be able to think reflectively about their own teaching, materials they bring to lesson, activities they apply, and assessment tools they use by looking at assessment result which later will help them evaluate their effectiveness and, if necessary, make changes in relevant areas. The following quote exemplifies the situation:

The teacher should be able to criticize himself/herself by looking at the results. The way s/he teaches, the methods s/he uses in the lesson, the assessment tool, the assessment method (perhaps the student knows but s/he cannot show his/her performance with that assessment method), the curriculum s/he follows... s/he should be able to think reflectively, evaluate them, and make changes if necessary. (Comparison pre-semi-structured interview, P4)

In addition, the participants stated that students should think reflectively on their own by looking at results. By thinking reflectively, students can ask by themselves this question "What can I do differently to improve the situation?" , find an answer to it and correct their own mistakes in their learning process. Also, participating instructors stating that they did not include reflective thinking in their assessments in the previous interviews stated that they believed that teachers, students and school administrators could increase the quality by reflecting on the assessment results. The following extract shows participant's opinion on how students should reflect on assessment results:

Grades are perceived as grades for students, of course, but at the same time, the student should take a lesson for his own development from that assessment result. By saying where I am wrong and where I have weaknesses, s/he should determine areas needing support and development and fill in the gap so hat s/he can improve. (Last pre-semi-structured interview, P2)

4.4.4. Teachers' Additional Developments Regarding Assessment

This theme consists of five sub-themes: teachers' developments in ethics, teachers' developments in administering exams, teachers' developments in characteristics for quality assessment, teachers' developments in what to consider to increase the quality of assessment, teachers' developments in relationships in the issues regarding assessment. Detailed analysis of each sub-theme is presented in the following sections.

4.4.4.1. Teachers' Developments in Ethics

Participants attended the RRD session on ethical issues in assessment and evaluation. Later, they wrote a reflection report, filled in a KWL chart and participated in the interview sessions. According to the results of the analysis of the data obtained as a result of this procedure, it has been revealed that the teachers have developed their cognition on ethical issues in assessment. The participants, who first thought of only grading when ethical issues were mentioned, stated that ethical issues should be considered at every stage of the assessment and evaluation procedure after the RRD sessions. Based on the statements of the participants, it can be said that they have developed themselves in knowledge and implementation of ethics in assessment, identifying the difference between ethical –unethical behaviors in assessment, ethical principles, influence(s) of ethical principles in assessment and association between ethics and reliability. Table 83 details the development of participants' cognition about ethical issues in assessment:

Table 83.

Teachers' Developments in Ethics

Code	f	Code Description
Development in knowledge and practice of ethical considerations in assessment	35	Development in participants' cognitions considering the knowledge and implementation of ethics.
Increased knowledge in un-ethical behaviour in assessment	14	Developments about how participants differ ethical behaviour from unethical ones.
Development in ethical principles	6	Development in ethical principles, what ethical principles are
Development in influence of ethics in assessment	2	Development in participants' cognitions about how ethical issues influence assessment practices.
Increasing reliability for ensuring ethics in assessment	1	Development in enhancing reliability in assessment.

When Table 83 is examined, it can be seen that the participants showed a significant development in their cognitions regarding ethical issues. Prior to the RRD sessions, participants associated ethical issues only with fairness and objectivity in grading. After the RRD sessions, they stated that ethical issues should actually be considered at every stage of assessment process. In addition, the participants, who stated that ethical issues are more comprehensive and important than they thought, stated that they would definitely pay attention to ethical issues at every stage of their following assessment practices. The following excerpt exemplifies this finding:

Actually, I did not know that ethical issues were such a comprehensive issue. I would just try to be ethical in grading. After the RRD sessions (suppose we are going to administer a written exam to the students), I learned that ethical issues should be taken into consideration at every stage. For example, I now know that ethics must be ensured in test preparation, test item writing, scoring, administering the tests, grading, announcing the test results and even in the decisions made based on the test results. I learned how to behave ethically at every stage I mentioned, and one of the elements that I will especially consider from now on will be ethical issues. (Participant reflection report, P3).

Moreover, knowledge level of the participants about which behaviour is ethical and which behaviour is unethical in assessment and evaluation has increased. For example, consciously applying a difficult exam to students in order to punish them, or asking a question from the subject which was not taught, or purposely assigning high scores to difficult questions and low scores to easy questions... They were among the behaviours that the participants sometimes reported to show. However, after the RRD sessions, the participants stated that they learned that these behaviours were unethical and would not do them again. The following excerpt exemplifies the participants' development of cognition about ethical and unethical behaviour:

Asking out-of-content questions is a violation of rights. The student's rights should not be violated. It is necessary to pay attention to the content validity. It is also unethical to give test questions in advance or to show questions similar to those that will appear in the exam in the classroom. I learned about ethics. In addition, we do not punish students by asking difficult questions. It is unethical to treat or ask questions in a way that will emotionally injure the student or make him/her feel bad. At the end of the RRD session, of course, I learned much more than these, and I will pay attention to all of them. (Participant reflection report, P2).

Last but not the least, the participants stated that they did not know the ethical principles in assessment before the RRD session, but after the reading, they had detailed information about these two principles such as 'Give no harm and Avoiding score pollution'. They reported being better aware of these principles and implementing them in their future assessment practices.

I learned two main principles in ethical matters. Avoid score pollution and do no harm. Actually, I somehow ensure them in my practices. Do no harm impressed me

*the most. We must protect the rights of the students. Our assessment practices should not harm the students emotionally and physically. We have to be very careful for this. I'll be more careful from now on to ensure this. In addition, **if we give the students the score they deserve, we will prevent score pollution.** (KWL charts, P4).*

4.4.4.2. Teachers' Developments in Administering Exams

Through this sub-theme, teachers' self-perceived development in their cognitions regarding administering exams is presented. Teachers reported that their development is related to administering exams to student, being a proctor during the exams, using suitable assessment technique or tool to increase reliability while administering exams. Table 84 provides a detailed investigation of this sub-theme.

Table 84.

Teachers' Developments in Administering Exams

Code	f	Code Description
Development in administering exams	15	Development in how to administer exams
Being proctor in administering exams	13	Development in being proctor during exams
Giving enough time and being reminder in administering exams	2	Development in how to behave appropriately during the exams
Using suitable assessment tool for ensuring reliability in administering exams	1	Development in how to enhance reliability in administering exams

When results are examined, it is seen that the participants have shown a great development in administering any exam to the students. The participants, who had previously only attempted to prevent students from cheating, now clearly stated that they have more control over what needs to be done before and during the exams. According to the participants, they have become more knowledgeable about suitability of physical environment of the classroom, testing the validity and reliability of an assessment tool, giving short, clear and understandable instructions to students during the exam, and the day or time of administering the exams. The following excerpts shed more light on the teachers' cognition development concerning administering exams.

*There are responsibilities of the teacher rather than the student. Before the exam, we **teachers are very nervous, we shout at students to make students feel that seriousness** that is to prevent students from cheating, we create a tense environment. I learned that **this is wrong**. On the contrary, **we need to be calm and moderate so as not to get stressed the students**. Students should also be prepared for the exam, this should be taken into account. **I learned that the dates, the days when we will apply***

the exams, should not be after the physical education classes, for example. I have to administer exams before or after lunch. These can also increase the inclusion of measurement error. We need good timing so that we can get reliable results. (Comparison pre-semi-structured interview, P3).

*I learned useful information about administering the exams. **Physical environment of the classroom, the measurement tool, the teacher, that is, the assessor and the student should be organized for administering exams.** The physical environment of the classroom should be suitable for the exam in terms of **heat, light and sound.** **The readiness level of the student** must be appropriate for the exam. **The teacher** must also be **physically and psychologically ready to administer the exam.** If he is not feeling well, he should postpone the exam. **The teacher should not get students stressed** during the exam. The assessment tool should be legible and understandable. **Ambiguity should not be included both in the statements in the exam and in the instructions given during the exam.** I learned that it is necessary to pay attention to these while applying the exams. I really didn't know. (Last post-semi-structured interview, P1).*

Another development of the participants related to administering exams is related to supervising that is observing test takers during the exam. After the RRD session, the participants had an opportunity to realize and correct many mistakes that they made. They learnt a word called proctor and learnt how to be an effective proctor. The participants stated that after the end of this study, they would not make the mistakes they made before and would be a good proctor during administering the exams. The following extract explicates the situation better:

*Another information that I learned about the subject is the word “**proctor**”. In other words, **the supervisor should not read books during the exam, and should not play on the phone. S/he should be a good observer without moving much.** It won't lead the students to get stressed, which is an ethical behaviour. It will help students to perform in the exam comfortably by creating a peaceful environment. (Comparison pre-semi-structured interview, P4).*

4.4.4.3. Teachers' Developments in Characteristics for Quality Assessment

During the interviews held after the RRD sessions, the participants were asked the question "How do you think an effective assessment should be?". While answering this

question, the participating instructors talked about self-perceived teacher competencies to achieve this goal. The findings about these teacher competencies that emerged as a result of the analysis were brought together under this sub-theme. One of these perceived competencies is being assessment literate. Based on the results, teachers should also be competent in using assessment results, knowledgeable in assessment techniques and language assessment principles, and ethical considerations. Detailed explanations of the sub-theme are presented in Table 85:

Table 85.

Teachers' Developments in Characteristics for Quality Assessment

Code	f	Code Description
Perceived teacher competency for assessment- assessment literacy	9	Development in participants cognitions regarding knowledge of assessment literatures
Perceived teacher competency for assessment technique knowledge	3	Development in participants cognitions regarding necessary competency for teachers to assess their students suitably
Perceived teacher competency for assessment- ethical considerations	3	Development in teachers knowledge base considering competency for teachers considering ethics in assessment
Perceived teacher competency in grading	3	Development in cognition of teachers considering grading
Perceived teacher competency for assessment-use of results	3	Development in participants' cognitions considering competency for teachers regarding use of assessment results.
Perceived teacher competency for assessment-principles	2	Development in participants' cognitions considering competency for assessment principles

First of all, the participants believed that for an effective assessment, teachers should have adequate knowledge level of assessment and evaluation. The participants, who claimed that there was a significant development in their cognitions about assessment after the RRD sessions they attended, stated that assessment was a much more important subject than they thought, and that teachers should have special knowledge about it. Moreover, they stated that teachers who have sufficient knowledge about assessment and evaluation could make an assessment that improves teaching and learning. The following extract illustrates this finding:

*Teachers should be assessment literate. Today, I've learnt this concept. The teacher's assessment knowledge should be sufficient. In other words, **the teacher should have a good grasp of the concepts we mentioned about assessment and evaluation.***

Teachers should have a control of the whole process. S/he must be knowledgeable.

Assessment should be done deliberately, not randomly. (Participant reflection report, P1)

In addition, according to the participants, teachers' assessment technique knowledge should be rich. Participants, whose assessment technique knowledge was previously limited to traditional ones, stated that they increased their level of knowledge significantly after RRD

sessions. They determined that they would incorporate various assessment techniques suitable for their goals in their future assessment practices. Based on their own developments, the participants stated that for an effective assessment and evaluation to be implemented, teachers must know all kinds of assessment techniques, both classical and contemporary, so that they can make valid assessment. The excerpt below exemplifies the situation:

*Teachers should always **update themselves**. Their assessment **technique knowledge level should be high in quality**. For instance, different technique may work for the student who cannot show himself/herself. The student should choose the appropriate assessment technique. Or the technique should be suitable for the assessment objectives. For example, speaking skill cannot be assessed with a written exam.*

*Teachers should **determine and apply the appropriate technique for the purpose set accordingly**. Therefore, for an effective assessment, the teacher should have adequate level of knowledge about assessment techniques. (Comparison pre semi-structured interview, P3).*

One of the self-perceived teacher competencies put forward by the participants is related to grading. The participants stated that they had improved themselves in grading after the RRD sessions. The participants, who claimed that they would give more defensible, evidence-based and valid grades in their follow-up grading practices, suggested that every teacher should have an effective knowledge of grading. That's why they can make valid and reliable decisions based on assessment results and have a chance to improve quality of teaching and learning. The following explanation taken from the interview clarifies this emerging finding:

*As teachers, we should know **what grades reflect**. Grades should reflect student's **actual level of competency**. Teachers should know why they give grades? As teachers, we should **not give grades as a way of punishment**. We should know the criteria for **defendable grading**. Now, after the sessions, I am much more informed about grading. Teacher should be well informed as well. (Participant reflection reports, P4).*

4.4.4.4. Teachers' Developments in What to Consider to Increase the Quality of Assessment

The findings obtained from the answers given by the participants to improve the quality of assessment process and for better and effective assessment practices were gathered under this current sub-theme. In order to increase the quality of any assessment, asking right questions, involving all stakeholders in the assessment process, being a teacher who is

constantly in pursuit of learning and renewing himself /herself, enhancing assessment triangulation and conducting action research are the ideas frequently expressed by the participants. Table 86 illustrates the detailed investigation of this sub-theme:

Table 86.

Teachers' Developments in What to Consider to Increase the Quality of Assessment

Code	f	Code Description
Conducting action research for improving quality of assessment	4	Development in participants' cognitions considering carrying out action research for increasing assessment quality.
Collaboration with stakeholders for improving quality of assess	4	Development in increasing quality of assessment through cooperation with stakeholders of assessment.
Asking right questions for improving quality of assessment	3	Development in teachers' cognitions considering asking right questions in assessment practices.
Multiple assessment for increasing quality of assessment	3	Development in ideas of how quality of assessment can be increased through multiple assessment.
Unbiased grades for improving quality of assessment	3	Development in participant's cognition considering giving grades, obtaining unbiased grades.
Being an updated teacher for increasing quality of assessment	1	Development in being open to new developments in language teaching along with assessment.
Grading separately for improving quality of the assessment	1	Development in participants' cognitions considering improving quality of assessment – grading separately

Participants stated that in order to increase the quality of assessment and evaluation, quality data should be collected and the right questions should be asked for this purpose. The participants, who stated that they did not do any research for the questions they prepared or chose before, stated that they understood the importance of being able to ask the right questions after the RRD sessions. By only asking right namely valid and reliable questions, teachers believe that they can derive quality assessment data from their practices and make sound decisions for putting learning and teaching in a better place.

I learned that asking questions is important. We said "quality data should be used in the interpretation of the assessment results and in making decisions based on the assessment results. For the use of this quality data, quality questions should be asked so that we can take appropriate and valid steps. (KWL charts, P3).

Later, as the participants mentioned about the interpretation of assessment results, they suggested that help from the necessary stakeholders should be appreciated throughout the entire assessment process. Therefore new ideas can easily come up through the exchange of ideas and as a result, more effective improvements can be achieved. They believe that quality of assessment will be improved by collaborative practices, such as asking a colleague to check the exam before administering it, or asking a colleague to re-grade the exam papers to ensure inter-rater reliability in the grading process. This result can be better understood looking at the excerpt:

*However, I got the point that **it would be better to collaborate with other teachers to improve the quality of assessment. By sharing ideas on assessment procedure, we can come up with better ideas and we can apply them in our actual practices.***

Consequently we can improve the quality of assessment. (Participant reflection report, P2).

Moreover, in order to increase the quality of assessment, participants expressed what they learnt as a result of RRD sessions related to grading. The participants, who stated previously that they tried to only be objective in the grading process, learnt the triangulation of assessment in the RRD session. Based on this, another way to increase the quality of assessment according to the participants is to assess students with several different assessment tools or techniques and to evaluate them holistically in the light of the data obtained from them. The excerpt explicates the current finding:

*We decided at the end of the session that **triangulation of data was important.***

*Through it, **more reliable results can be obtained.** Decisions made based on those **results are more valid, so if we want to increase quality, I think we should include more than one assessment tool in the process.** (Participant reflection report, P4).*

Finally, the participants touched on action research. Participants, who had not known what action research was before the sessions, first learnt what action research is and its steps in the RRD session. As a result, the participants put forward their ideas by connecting the conducting action research to how the quality in assessment and evaluation could be increased. According to the participants, a problem related to the applications in the assessment and evaluation process will be determined and then an action research will be carried out on it, and the problem will be solved and ultimately the quality will increase.

*In addition to these, **teachers, students and administrators can conduct a study in order to increase quality. This is the action research.** I've heard this type of research but I did not have extensive knowledge about it. I searched it, I read some and I am a bit illustrated. Since I was graduated of ELL, I did not take assessment course. For that reason, these reading and discussion sessions are so informative for me. They work really best for me. From the readings I've learnt that action research has some steps to be taken. First of all, a problem or a subject should be determined. Then they should collect and analyze the data. At the end, results are reported and much informed actions can be taken. **Therefore, quality of education, assessment in our***

case, can be developed. Problems can be solved and effective decisions about teachers' teaching practice, students' learning or effectiveness of assessment practices can be made. May be as instructors, we can conduct an action research with our students and administrators in order to obtain useful data then we can make decisions about our students, improvement of our school of foreign languages and our teaching practices by taking the obtained data into account. Therefore we can use the data to ignite change. (Participant reflection report, P3).

4.4.4.5. Teachers' Developments in Relationships in the Issues Regarding Assessment

In this sub-theme, there are research findings that refer to relationships that participants have inferred among definitions about assessment. These emerging findings were found to be relationship among assessment, teaching and learning. Also, findings reveal teachers' development in their cognition regarding relationship between valid assessment results and quality instruction, and association between testing and assessment Table 87 presents detailed investigation of this sub-theme:

Table 87.

Teachers' Developments in Relationships in the Issues Regarding Assessment

Code	f	Code Description
Interconnectedness of assessment and teaching learning	7	Development in participants' knowledge regarding relationship among learning, teaching and assessment and how they influence one another mutually.
Relation of valid assessment results to teaching quality	5	Development in participants' cognitions regarding how quality of teaching and learning is influenced by valid assessment results.
Development in relation between testing and assessment	3	Development in participants' knowledge level considering relationship between testing and assessment, how they are related to one another.

Based on the results, it can be concluded that the participants have established a link among assessment, teaching and learning. The participants, who could not detect this relationship before the RRD sessions, see assessment, learning and teaching as elements that are in constant interaction with each other as their level of knowledge increases as a result of the RRD sessions. According to the participants, the quality of learning and teaching depends on the quality of assessment and evaluation. Only with valid and reliable results, valid inferences about students' learning and teachers' teaching can be made. The following excerpt exemplifies the participants' views:

This is not just about only students. Indeed, after these sessions, I realized that assessment, learning and teaching are intertwined and affect each other. In other words, if an assessment process is not conducted with a valid tool and in a reliable

*way, the result obtained will be invalid. According to this result, maybe the teacher will change his/her teaching. Maybe the student will change his/her learning by looking at that result. **If these are done with an invalid assessment result that does not reflect the truth, it will be more harmful than beneficial.** As you can see, assessment affected both teaching and learning. (Comparison pre-semi-structured interview, P1).*

In addition, by understanding the distinction between assessment and testing, the participants came to the conclusion that assessment is an action beyond testing and applying tests to students, as exemplified below.

***Every test is an assessment, but not every assessment is a test.** Yes, that's why after the sessions, I realized that assessment is more than administering pen and paper exams. For example, assessment includes every type of gathering information (Participant reflection report, P3).*

4.4.5. Teachers' Developments in Introductory Issues about Assessment

As a result of the analysis of the obtained qualitative data, the fifth theme of the research findings, teachers' developments in introductory issues about assessment, emerged. This theme is related to the developed definitions and perceptions in the cognitions of the participants about basic concepts related to assessment, the purposes of assessment, and assessment types, especially formative and summative assessment. These three emerging sub-themes are examined in detail one by one in the following sections.

4.4.5.1. Teachers' Developed Perceptions and Definitions of Assessment

For the current sub-theme, seven codes that were most frequently mentioned and considered important by the researcher were included in the Table 88. These codes are the definitions of the most frequently encountered concepts related to assessment and evaluation. Table 88 details this sub-theme:

Table 88.

Teachers' Developed Perceptions and Definitions of Assessment

Code	f	Code Description
Perception of assessment	47	Development in participants' perceptions considering assessment as a whole
Development in difference between measurement and assessment	12	Development in participants' cognition considering definitions of measurement and assessment.
Increase in knowledge in valid assessment results	8	Developed participants subjective definitions considering valid assessment result.
Increased knowledge in definition of evaluation	7	Development in the definition of evaluation.
Development in difference between assessment and evaluation	4	Development in difference between assessment and evaluation, how participants define assessment and evaluation separately.
Development in perception of effective assessment	3	Development how participants perceive effective assessment. How an effective assessment should be.
Increased knowledge in definitions of in /formal assessment	4	Development in definitions of formal and informal assessment.

As a result of the RRD sessions attended by the participants, it was revealed that their perceptions about assessment improved significantly. In previous interviews, participants matched assessment and evaluation with pencil-and-paper tests. Some participants also perceived assessment as merely quantifying students' performance in order to determine their learning levels. In general, it can be said that the assessment perception of the participants before the RRD session was to subject students to exams, especially pencil and paper exams, in order to determine their learning levels. After ten RRD sessions they attended, all of the participants stated many times that there was a positive development in their perceptions about assessment. They also indicated that, according to the participants, assessment is a broad procedure beyond including students in the exam. One of the participants even described his/her assessment perception before the RRD session as a drop in the ocean and the assessment as the ocean. Thus, it can be said that the participants' perception and knowledge of assessment before the RRD session was limited. To highlight it again, the participants proved the development of their cognition of assessment, both with the KWL chart and reflection reports they prepared, and with the answers they gave to the questions in 3 separate individual semi-structured interviews they attended after the RRD session. The following excerpts illustrate this development in the participants:

I have something to add. With Assessment, we can follow the progress of the students and see their strengths and weaknesses. We can also place the students. We can include them in a group. Then, when we look at ourselves as teachers, we can also make changes in our own knowledge and in-class practices. We can get feedback from assessment. If we want, we can change the curriculum, we can change our

method. *We can obtain more reliable and valid data by determining our objectives and choosing an assessment tool related to it. I learned that assessment is a very broad concept, that it is not just about the student, it also contributes to the teacher. After the sessions these are the ones I added to my level of knowledge. Of course, there are some that I can't remember. (Comparison pre-semi-structured interview, P3).*

*Previously, I used to give the paper and pen to students. If the student has answered, it is an assessment. But I learnt that it wasn't. I learned that **assessment encompasses all processes such as grading, evaluation, question preparation, setting objectives, considering ethical issues, etc.** (Comparison pre-semi-structured interview, P2).*

*I used to know assessment as **giving and applying tests to see student performance, then scoring that test and then making a decision considering those scores**, but I learned that **assessment is a very broad concept. It also includes testing, that is, from the very beginning, from preparing a question to applying it, sharing the results, analyzing them, making decisions based on the results and making the necessary changes...** Assessment has many stakeholders, many dynamics and all interact with each other. (Comparison pre-semi-structured interview, P1).*

*Yes, there **has been a progress**. Before I started the sessions, I used to perceive the assessment as just a written exam. I said "Assessment is giving grades to the students". But after attending these sessions, I learned that **assessment is a different process than giving grades. In this process, I learned that it is necessary to gather information from many sources in order to understand the student in detail and to help him/her develop.** Later, I learned that it is a process that we determine and apply how we should analyze the information we have obtained or how we should use the information in order to improve the student, teaching and school. In other words, I learned that **assessment is a process that covers not only grading in written exams, but also different processes.** (Last pre-semi structured interview, P4)*

Since the main focus of the study was to show the development, if any, of the participants' cognitions related to assessment, the development stated by each participant was exemplified one by one.

4.4.5.2. Teachers' Developments Regarding Assessment Type

This sub-theme includes the findings of the development in the cognition of the participants about assessment types. Participants talked about formative, summative, criterion and norm-referenced assessment in the data they presented about assessment types. A detailed explanation of the sub-theme consisting of findings referring to the development in the cognitions of the teachers about assessment types is given in Table 89:

Table 89.

Teachers' Developments Regarding Assessment Type

Code	f	Code Description
Development in knowledge/ perception of formative assessment	27	Participant reported to develop himself or herself considering formative assessment. Perceived increased knowledge in formative assessment
Development in knowledge/perception of summative assessment	18	Participant reported to develop himself or herself considering summative assessment.
Development in implementation of formative assessment	15	Development in implementation of formative assessment in actual assessment practices.
Development in criterion-norm referenced assessment	3	Development in assessment type, i.e. criterion – norm referenced assessment
Integration of both summative and formative assessment	3	Development in incorporation of summative and formative assessment into actual classroom assessment practices by integrating them simultaneously.
Ensuring formative assessment through traditional techniques	2	Enhancing formative assessment through using traditional assessment techniques. Development in implementation of traditional techniques for enhancing formative assessment.

It can be seen that there is an improvement in knowledge levels of the participants about formative assessment. Participants knew beforehand about formative assessment that it was only process-oriented. In one of the interviews before the RRD sessions, it was said that the formative assessment and the formal assessment were the same. There was even a participant who claimed that the formative assessment was an alternative assessment technique. However, after the RRD sessions, the participants learnt what formative assessment is and its advantages. This emerging finding illustrates the situation:

*Before the reading and discussion, I used to think that **formative assessment equals to alternative techniques**. Thanks to this reflective reading discussion session, now **I can clearly define what a formative assessment really is**. Thanks to this session, I can say that I should adopt **formative assessment because it is an assessment for learning**. But I also know that **summative assessment is not a sin** or formative assessment is not a totally advantageous. (Participant reflection report, P3).*

In addition, the participants talked about how they can implement formative assessment in their own actual assessment practices as a result of RRD sessions. In fact, it was previously

found that the teachers do not need to use only contemporary assessment techniques to ensure formative assessment, but they could also provide formative assessment with classical assessment methods. The participants, who stated that the most important thing in this issue was the willingness and knowledge of the teacher, claimed that with appropriate planning and good management of the process, they could also perform formative assessment on their own, even by using classical assessment techniques. Excerpts are given to illustrate:

Last but not least, before these sessions I used to know that formative teachers do not use traditional assessment techniques. They use contemporary ones. Thanks to this session, I've learnt that formative teachers who want to have diagnostic information about their students' can use multiple choice test which is paper and pencil test in order to see students' levels in the process. (Participant reflection report, P3).

4.4.5.3. Teachers' Development of Assessment Purposes

In this sub-theme, there are findings that refer to the development in participants' cognitions about assessment purposes as a result of RRD sessions. Previously, almost all of the participants claimed that they assessed to determine whether students learnt the subjects taught to them and to make decisions about student success. After the RRD sessions, a self-perceived increase in participants' assessment purposes emerged as a result of the analysis. Table 90 details the sub-theme:

Table 90.

Teachers' Development of Assessment Purposes

Code	f	Code Description
Assessment for developing or changing curriculum	2	Participants reported to assess their student in order to check the efficacy of the curriculum and make necessary changes.
Assessment for evaluating teaching efficacy	2	Development in assessment purpose. Assessing students for checking effectiveness of teachers' teaching practices.
Assessment for identifying students' strengths and weaknesses	2	Development in assessment purpose. Assessing students in order to shed light on students' strengths and weaknesses.
Assessment for school improvement	2	Development in purpose of assessment. Assessing students in order to improve school standards.

In the interviews held before the RRD sessions, the participants stated that they were assessing to test effectiveness of their teaching to some extent. However, they did not state that they assessed for school improvement, updating or changing the curriculum, or clearly identifying the weaknesses of the students and planning their instruction accordingly. After the RRD sessions, participants varied their assessment purposes. For example, the participants stated that they would assess in order to test the effectiveness of the curriculum used in reaching the objectives and to change or adapt the parts that they deem necessary. In addition,

it was found that they would assess the students' expectations and solution suggestions in order to ensure school development. Based on these findings, it can be said that the cognition levels of the participants related to the assessment purposes have improved.

4.4.6. Teachers' Developments in Assessment Principles

The results that emerged from the statements of the participants and referred to their cognition development related to assessment principles were gathered under the title of this theme. The theme consists of three sub-themes. Reliability, which is one of the assessment principles, is handled as a separate theme. In addition, practicality and authenticity, which are among the assessment principles, were found to be self-perceived strengths of the participants according to the analysis of the interview data obtained before the RRD sessions. Since the purpose of this research question, the 3rd one, was to reveal the perceived development of the participants' cognitions after RRD sessions, washback and validity were included in this theme separately as a sub-theme. In addition to these, the findings of the consistency were also categorized and joined this assessment principles theme. The sub-themes referring to the results related to teachers' perceived development in their cognition concerning assessment principles are explained in the following sections accompanied by separate tables.

4.4.6.1. Teachers' Developments in Washback

During the study, the participants read about the washback and then participated in the discussion session. They also answered questions on the subject during the interview sessions. The analysis of the participants' answers shows that they developed their cognitions about the washback. Participants, who previously knew nothing about washback, now learned what washback is, its importance, how to enhance it, and possible ways to produce washback. Table 91 presents the participants' cognition developments regarding washback in detail.

Table 91.

Teachers' Developments in Washback

Code	f	Code Description
Increased knowledge in washback	31	Development in knowledge of washback, what it means, what it provides
Enhancing washback through sharing assessment results	26	Development in enhancing washback while sharing assessment results.
Washback through using a rubric	13	Development in enhancing washback as a result of using a rubric.
Washback through portfolio	6	Development in: enhancing washback by employing portfolio.
Washback through alternatives	6	Development in creating washback through alternatives.
Washback through self-assessment	3	Development in enhancing washback with the help of self-assessment.
Washback through classroom observation	2	Development in enhancing washback with the help of classroom observation.
Ensuring washback through formative assessment	5	Development in ways of enhancing washback via formative assessment.

As mentioned before, it was revealed as a result of the analysis that participants had a great deal of weaknesses related to washback before the RRD sessions. EFL instructors, who read and discussed about the subject, made a distinction between washback and feedback after the sessions, and learnt that washback can be enhanced through constructive feedback. They also alleged that washback is a very important tool to increase the competency levels of students, and that after the assessment practices or after explaining the results to the students, washback must be enhanced, otherwise it will not have a positive effect in any way. The cognition development of the participants about washback is exemplified by the excerpt:

The effect of the exam on the student is washback. I didn't know this, but I learned. The exam should have a positive impact on students. Exams and their results should enhance positive washback for student. It should show the way... The exam should allow the student to see his/her weaknesses so that the student can make progress on his/her own with constructive feedback from the teacher. At the same time, exam and its results should ensure that the student sees his/her strengths in order to be motivated. I learned that washback is a tool for both guiding and motivating students. Now I will take care that my exams create a positive washback on my students.
(Comparison pre-semi-structured interview, P1).

Next, the participants learnt that they could enhance the positive washback while announcing the results especially to the students. Before the RRD sessions, the participant instructors who explained assessment results through the system stated that they did not give any feedback to the students considering the results, but they realized their mistakes they

made after the RRD sessions. Participants indicated that continuing teaching without doing anything after announcing the assessment results did not cause any improvement for the students. According to them, informing the students about strong and weak parts of their performance and creating a possible action plan with the students in order to compensate for their deficiencies would be beneficial for them thus a positive washback could be ensured and students could be facilitate towards desired direction.

*What I like most about that week is washback. I must admit that I almost never implemented washback after an exam. **I knew that it means giving feedback to the students after analysing their papers.** However, all I have done just announced their grades. I didn't assess any paper with an analytic eye. **During the discussion I realized that washback is more than a slight feedback or just letting the students know their grades.** For my next exams am I going to own a proper washback? Well, definitely I will be well aware of its aim, and I will assess the students' test items with this responsibility. I will try to increase a private washback for every student about their weak and strong parts in learning. **I will also increase a general washback addressing all students with the aim of informing them about their performances; what they did well, in what parts they should be more careful, and for next time how they should prepare.** (Participant reflection report, P2).*

One of the self-perceived cognition developments of the participants is related to the creation of washback for students and teachers through alternative assessment. As a result of the RRD sessions, the participants who improved themselves on both washback and alternative assessment concluded that washback could be ensured through alternative assessment. To illustrate, the participants stated that alternative assessment is advantageous in terms of providing construct validity, considering individual differences and even in terms of its capacity to present individualized assessment. Also, they added by stating that via applying alternative assessment, each student can be reached effectively and positive washback through feedback could be ensured. This situation is presented through the participants' statements:

*Also **alternative assessment** provides information about students' strengths and weaknesses. Also, since it **focuses on the process, we can detect student's weaknesses immediately in time and we can take some precautions** by warning students or facilitating their learning. As teacher, thanks to alternative assessment, **we can give***

feedback to students and help them correct their errors before it is too late. (KWL chart, P3).

4.4.6.2. Teachers' Developments in Validity

The findings on development of the participants' cognitions related to validity have been gathered under this sub-theme. Before the RRD sessions, the participants knew only about content validity. They could not fully define validity, they did not know the types of validity. It was revealed that the participants, who were unaware of the role of validity in the assessment process, had weaknesses about validity. After the RRD sessions, the participants stated that they had development on validity. Table 92 which is seen below includes the findings referring to the self-perceived development of the participants concerning knowledge in validity especially construct validity, relationship between validity and reliability, and valid assessment result:

Table 92.

Teachers' Developments in Validity

Code	f	Code Description
Increased knowledge in validity	24	Development in the knowledge of "what is validity in general?"
Development in knowledge of construct validity	13	Development in participants' knowledge of construct validity
Development in relation between validity and reliability	10	Development in connection between validity and reliability
Valid assessment result	10	Development in knowledge regarding valid assessment results.
Ensuring validity by Obtaining reliable results	3	Development about how is validity enhanced as a result of getting reliable results.
Valid assessment tool	2	Development in the conception of "valid assessment tool"
Difference between content and construct validity	1	Development in difference between validity types: content validity and construct validity.

As seen from the results, participants' most frequently development is related to what is validity in general. Before the RRD sessions, the participants could not fully define validity, and when they were asked for their ideas about validity, all of the participants talked about content validity. After the RRD sessions, the participants effectively defined validity. They explained types of validity by giving examples of each. Emphasizing the importance of validity for a quality assessment, the participants stated that they are now well informed about the issue and that they will ensure validity in their assessment practices after the sessions. The excerpt exemplifies this finding:

*Validity... Can we assess exactly what we want to assess, what we aim for, with the assessment tool or technique we have adopted. There are kinds of validity, such as face validity, construct validity. **I learned construct validity.** For example, let's say I'm going to assess the student's ability to generate word stress. I cannot assess this with written, pen-and-paper tests. **Whatever construct I want to assess, there should be an assessment tool that I can assess exactly.** Likewise, an assessment result with a high construct validity is a result obtained in this way. There are other types of validity such as content validity and predictive validity. **Now I know more about validity.***
(Comparison pre-semi-structured interview, P1).

In addition, it was stated that the teachers learnt the relationship between reliability and validity by making inferences. Participants who did not have detailed information about both validity and reliability before the RRD sessions showed development after the sessions they attended. Participants argued that validity and reliability are interconnected, and it is important to ensure reliability in ensuring validity. The excerpt, which exemplifies the development in the cognition of the participants who stated that not every reliable assessment result or tool can be valid at the same time, is given below:

***Not every reliable result is valid, but all valid results are reliable.** To ensure validity, we must first establish reliability. I have learnt this mutual interconnectedness.*
(Comparison pre-semi-structured interview, P3).

Also, participants were asked questions about the valid assessment result during the interviews. Based on their answers to the questions, it can be said that the participants have cognitive developments related to the subject. Instructors stated that they learned what a valid assessment result means, how a valid assessment result can be obtained, and what kind of advantages are provided by valid results. In addition to these, they reported that they would pay necessary attention to the validity of the assessment result they have obtained through their own assessment practices.

*Valid or invalid assessment result... We discussed this, we said "the assessment tool we will use, for example our exam, should be a valid. We must administer it in a reliable way. The content validity of our exam should be high, the construct validity should be high, the face validity should be high. **When we apply such a valid test to students in a reliable way, our results will be valid results.** Students' current competency levels will be realistically revealed with a valid result. (Last semi-structured interview, P4).*

4.4.6.3. Teachers' Developments in Consistency

In this sub-theme, there are findings that refer to the cognition developments of the EFL instructors regarding consistency. This perceived development includes increase in teachers' knowledge of consistency, increase in enhancing consistency in assessment practices respectively. Table 93 details this sub-theme:

Table 93.

Teachers' Developments in Consistency

Code	f	Code Description	Example
Increased knowledge in consistency	12	Development in participants' cognitions considering consistency such as its meaning, its role in assessment, its contribution	Consistency in assessment results should be high. Consistency positively influences reliability. If we enhance consistency, we will ensure reliability. Consistency affects reliability, reliability affects validity, then the quality of the assessment increases. (Last post-semi structured interview, P3).
Increased knowledge in enhancing consistency	12	Development in participants' cognitions considering ways of enhancing consistency in assessment.	I have learned that consistency can be done in two ways such as intra-rater and inter-rater. I have realized that grading is much more difficult than we think. For example, we give 56 to a student performance. Does his / her performance really worth. 56? We should clearly explain and defend why we gave that grade. (KWL Chart, P1).

When Table 93 is examined, it can be seen that participants have cognitive development on both what consistency is and how to ensure consistency in assessment practices. Before the RRD sessions, some of the participants did not know consistency while some of them were found to have wrong knowledge about consistency. After the RRD sessions, the participants compensated for these weaknesses. The participants, who claimed to have learnt about consistency by referring to the consistency-reliability relationship, indicated that they would ensure consistency in assessment results with intra-rater and inter-rater reliability.

4.4.7. Teachers' Developments in Grading and Scoring

Under this theme, findings referring to the teachers' cognition developments concerning grading and scoring were grouped and included. According to the analysis result, the theme consists of four sub-themes. They are listed as teachers' developments in grading in general, teachers' developments in reliability in grading, teachers' additional developments in grading and scoring, teachers' developments of ethics in grading depending on their frequency of citation. In the following sections, each sub-theme is explained in detail with tables.

4.4.7.1. Teachers' Developments in Grading in General

Before the RRD sessions, the participants perceived grading as merely determining the students' performance by numbers. When asked questions about grading, they talked about grading written exam papers and being fair while doing this. After the RRD sessions, it can be said that the cognitions of the participating instructors about the grading improved. Based on the results, teachers' perceived development concerning grading was found to be related to how to grade, using a rubric in grading, washback through grading, perception of the nature of grades, and the importance of giving defensible grades. Codes constituting this emerging sub-theme are presented in Table 94:

Table 94.

Teachers' Developments in Grading in General

Code	f	Code Description
Development in how to grade	36	Development in participants' cognitions considering grading as an action in assessment
Using rubric in grading	7	Development in grading, using rubric in grading, its advantages and how it is materialized.
Washback through grading	7	Development in participants' cognitions about how to provide with washback through grading.
Nature of grades	4	Development in the nature of grades, what it accounts for, and how it should be.
Giving defensible grades	2	Development in defensible grades and their relation to reliability and validity
Not to punish students in giving grades	1	Development in grading, the misconception of grading which is grading for giving punishment to students

First of all, participants stated that they had knowledge about how grading should be done. According to results, they learnt specific rules of grading, such as specifying grading philosophy along with informing it to the students in advance with certain grading criteria, and that they would achieve grading in a more systematic way in their following assessment practices. The participants, who reported to purposely grade by aiming for the student to get a high grade in the exams, learnt after the RRD session that the main purpose was not to help students get high grades, that this was a completely wrong practice.

In addition, the participants stated that while grading student's exam paper before the RRD session, they used to give higher or lesser grades by considering behaviour of the student throughout the term and frequency of their attendance. After the RRD session, the participants stated that this behaviour was definitely not appropriate and that they would grade them separately when they wanted to include the student's participation in the grading practices. In other words, they stated that they learnt that exam performance should be graded

separately, behaviour and course participation should be graded separately. Based on these, it can be said that there is a development related to grading in the cognitions of the participants as a result of RRD sessions. The result is shown in the excerpts:

The aim in grading is not to ensure that the student gets high grades, but to determine the current competency levels of the student. Determining whether the student has reached the student objective or not... so my previous attitude was completely wrong. (Comparison pre-semi-structured interview, P1).

*First of all, I will create grading philosophy, **determine grading criteria**. That is, how students' performances will be assessed. I will specify them in detail, and **I will share them at the beginning of the semester with my students**. (Participant reflection reports, P3).*

Moreover, participants were asked questions about grading and rubrics in the interviews before the RRD sessions. In the answers given, it was revealed that the participants did not use a rubric in grading. The participants, who stated that they were informed about both rubric and grading after the RRD sessions, stated that it was necessary to use a rubric in order to prevent inclusion of measurement errors in grading and to give more evidence-based grades. The following excerpt better explains the situation:

*As teachers, we have to explain our students why we gave that grade to students. **Our grades should be to the point that we can make effective decisions**. We should also justify our grades. Actually, I did not know it and I didn't do this. For this reason, we should **prepare a rubric or a note taking system** in order to convey the basis of our conclusions to our students. I didn't know this. (KWL chart, P3).*

Last but not least, participants stated that they sometimes graded students in order to punish them, to warn them harshly, to lead them to have a bad experience. They stated that they deliberately punished the students by giving less grades than they deserved. However, the participants, who understood that the purpose of the grading after the RRD session was neither to punish nor reward the students, stated that they would not punish their students with grades from then on, but would give whatever grades they deserved.

*Sometimes, I apply **grades as a weapon to protect my authority and punish students**. Our discussion has clearly revealed that it is not professional. The assessments **should not aim to punish the students; on the contrary, while assessing the students I should pay attention to***

the students' rights and give the grade they deserve. From now on, grades won't be a means of punishment. I won't do it again. (Participant reflection report, P2).

4.3.7.2. Teachers' Developments in Reliability in Grading

Before the RRD sessions, the participants were asked how they ensured reliability in grading. In the answers given, the participants stated that they ensured reliability in grading by acting objectively. After the RRD sessions, it was found that participants showed development in this subject. As a result of the analysis, the self-perceived developments of the participants about ensuring reliability in grading were collected under this sub-theme which is detailed in the following Table 95:

Table 95.

Teachers' Developments in Reliability in Grading

Code	f	Code Description
Ensuring reliability by using a rubric in grading	15	Development in ways of ensuring reliability in grading- using a rubric or a checklist
Considering rater reliability in ensuring reliability in grading	11	Development in ways of increasing reliability in grading- rater reliability
Assessing in multiple ways for reliability in grading	8	Development in ways of ensuring reliability in grading: assessing through multiple ways, techniques.
Setting assessment criteria for ensuring reliability in grading	5	Development in cognitions of grading: setting clear and attainable objectives.
Ignoring personal feelings for ensuring reliability in grading	3	Development in participants' cognitions considering grading: in order to get objective grades, forget about personal relations

We can say that, the participants most frequently stated that they have improved themselves in using rubric in grading. Before the RRD sessions, they used to employ an answer key with ensure reliability while grading. After the RRD sessions, the participants showed development in both preparation and use of the rubric. In addition, the participants, who learnt that it is important to use a rubric while grading, stated that by using it, they would both give a defensible grade and ensure reliability by preventing measurement errors. Explanations below demonstrate the emerging result:

I have to pay attention to many things during grading. In order to ensure reliability, I definitely need an answer key with me. Let's assume that I will have conduct a performance assessment in the speaking class, for example. This time I need to use something like a rubric. Another way to ensure reliability was to use rubrics; I didn't know that. By looking at the rubric, I can give defensible grades and prevent measurement error. (Last pre-semi-structured interview, P4).

It was also revealed that the participants also improved themselves on rater reliability thanks to RRD sessions. They did not pay attention to rater reliability during grading process before the sessions. After RRD sessions, they stated that they would not share the results with the students immediately after the grading, and that they would share the results with the students after providing consistency by grading them again after a certain period of time or asking a colleague to grade. Thus, they claimed that they would increase reliability through rater reliability during the grading process. This finding is seen in the statements obtained from one of the participants' statements:

Thanks to inter-rater and intra-rater, I ensure consistency in grading. I read the exam again after a certain period of time, this is intra-rater or I ask a colleague to read the exam. This is the inter-rater. In both cases, I compare the results with mine. If the consistency is ensured, reliability is increased. (Last post-semi-structured interview, P3).

In addition, the participants learnt about triangulation of grading. In other words, they learnt that they should assess students with some different tools before announcing the grades to the students, and then announce the results to the learners after grading them all. Assessing students in multiple ways then grading them accordingly is one of the self-perceived developments related to increasing reliability in grading. The excerpt below illustrates the situation:

Also, I have learned that in order to make our grading more reliable, we should not rely on only one way of assessment. I think, it is triangulation of assessment. I have learnt it. We discussed it as well. Before announcing final grade to the stakeholder, as teacher, we should assess our students from multiple ways. We should grade each of them separately then we can determine last grade. Therefore, we can enhance the reliability of grades. (KWL chart, P2).

4.4.7.2. Teachers' Additional Developments in Grading and Scoring

In this sub-theme, there are research findings that refer to cognition development of participants regarding scoring and grading. Table 96 details the four codes that make up this sub-theme such as teachers' development in how to score, difference between scoring and grading, and ensuring validity.

Table 96.

Teachers' Additional Developments in Grading and Scoring

Code	f	Code Description
Development in how to score	12	Developments in participants' cognitions considering scoring as an action in assessment.
Difference between scoring and grading	10	Development in participants' cognitions regarding the difference between grading and scoring, specific characteristics of each of them and how each of them contributes to assessment process
Ensuring validity in scoring	3	Development in enhancing validity in scoring.

According to the results, there is a positive development in cognitions of participants about scoring. Before the RRD sessions, most of the participants stated that they scored by targeting the students to get high grades. In other words, higher scores were assigned on easy questions that anyone can answer correctly. They also stated that they assigned low scores to questions that are difficult and that everyone might answer incorrectly. It was also found that the participants scored the opposite in order to punish the students. It can be said that after the RRD sessions, the participants were informed about scoring. After that, they stated that they would assign higher scores to the questions that assess whether important objectives were acquired or not, to the questions that assess whether the subjects they focused on in the lesson were learned or not. In addition, it was revealed as a result of the analysis that they learnt considering Bloom's taxonomy during scoring, that is, they should assign high scores to questions in the analysis and synthesis stage and lower scores to questions that require recall of the information. The excerpt is exemplified the finding as below:

During scoring, I will assign a higher score to the important questions, that is, to the questions whose subject was focus in teaching. If the question has a high level of item discrimination level, also, I will give a high score. If the question assess whether the objectives that I consider important have been attained or not, I will assign it a high score. For example, we learned that a very difficult question is also not distinctive. Therefore, I will assign a higher score to the questions with a high item discrimination level, not to the very difficult question. (Comparison pre-semi-structured interview, P3).

Moreover, it was revealed in the interviews held before the sessions that the participants did not know the difference between grading and scoring at the definition level. However, after the RRD sessions, the participants learnt that scoring and grading are separate concepts. The following statements detail the emerging result:

I learned the difference between scoring and grading. Scoring is assigning scores namely points to the exam questions before applying the exam, while grading is giving grades namely points to the students' answers after the exam, taking into account the correctness or incorrectness of their answers to the questions. I used to think that both are the same, that is, to grade students' answers after taking the exam. But now, I know about it. (Last post-semi-structured interview, P1).

In addition to these, participants' self-perceived developments related to developing valid grading procedures were revealed as a result of the analysis. According to the participants, a grading procedure applied by trying to ensure reliability also contributes to validity. They stated that they can create a valid grading procedure while grading exams results obtained by using assessment tools that can precisely assess the target behaviors they wanted to assess and ensuring reliability in this process.

If I grade the assessment result, which I obtained by providing validity and reliability, by trying to ensure reliability and consistency, I also ensure validity. As a result, valid grading procedure occurs. (Last pre-semi-structured interview, P3).

4.4.7.3. Teachers' Developments of Ethics in Grading

The participants, who stated that they did not have sufficient knowledge about ethical issues before the RRD sessions, and that they only provided ethics by acting objectively in grading, were informed about ethical issues after the sessions and stated that they would consider them in grading. Table 97 contains the findings referring to participants' cognition development about enhancing ethics in grading:

Table 97.

Teachers' Developments of Ethics in Grading

Code	f	Code Description
Using a rubric for ensuring ethics in grading	7	Development in ensuring ethics in grading by employing a rubric.
Blind eye for ensuring ethics in grading	7	Development in ensuring ethics in grading by applying the blind eye technique
Ignoring personal relations in ensuring ethics in grading	4	Development in ensuring ethics in grading by forgetting personal relations between teacher and students

According to the result, the participants will first use rubric to ensure ethics in grading. Participants who did not use rubrics before the RRD sessions learnt benefits of using rubric

after the sessions. They stated that they would ensure ethics in grading without being unfair by giving everyone the grade they deserve through using rubrics. This excerpt below exemplifies the situation:

*I've read about ethical issues. Ethics is a much broader issue than I thought. Ethics in grading is one of the sub-topics included in this topic. I learned that **using rubrics contributes to reliability, washback, and ethics in grading**. I can **provide reliability using rubrics**. **This increases ethics**. Because I give everyone the grade they deserve, measurement error is minimized. (KWL charts, P1)*

In addition, the participants mentioned blind eye technique while talking about ethics in grading. Before the RRD sessions, the instructors stated that they included the blind eye technique in the grading processes, but they did not know that this technique increased the ethics in grading. After the RRD sessions, the participants stated that with this technique, they could increase both objectivity and fairness, as a result, it also ensures ethics in grading process. The following statements might enhance the implementation of this finding referring to the cognition development of the participants related to ensuring ethics in grading with the blind eye technique:

***In order to be ethical, it is necessary to close the names of the students on the exam papers and give grades while grading the exams in order not to make mistakes in the assessment**. This is one of the ethical behaviors in assessment. I learned that by doing this, ethics can be achieved in grading. (Last post-semi-structured interview, P4).*

In addition to the findings above, the participants stated that no prejudice or compassion should be shown towards students while grading, and that the teachers should be free of bias. Additionally, they claimed that only teachers who give grades without personal feelings and opinions can ensure ethics in this process.

4.4.8. Teachers' Developments in Reliability

Before the RRD sessions, the participants stated that they only ensured reliability by preventing students from cheating during exams. After the RRD sessions, it was revealed that there was a development in the cognitions of the teachers regarding reliability. Findings about cognitive development of the participants related to the subject were gathered under this theme. The theme consists of two sub-themes. They are teachers' developments in ensuring reliability in administering exams and teachers' developments in factors influencing reliability

in assessment respectively. In the following sections, each sub-theme is explained in detail with the tables.

4.4.8.1. Teachers' Developments in Ensuring Reliability in Administering Exams

Before the RRD sessions, participants were asked questions about reliability and administering exams. According to the analysis of their answers, it was revealed that participants simply prevented students from cheating to ensure reliability when administering the exams. However, analysis of relevant data obtained from participants after RRD sessions revealed that they learned several ways to ensure reliability when administering exams such as arranging the physical context of the class, determining the suitable time for assessment, preparing comprehensible tests, giving clear instructions during the exams and taking students' readiness for exams. Table 98 details this sub-theme:

Table 98.

Teachers' Developments in Ensuring Reliability in Administering Exams

Code	f	Code Description
Considering physical context for ensuring reliability in assess	21	Development in administering exams, enhancing reliability by taking the physical context of assessment into consideration.
Considering assessment time for enhancing reliability in assess	8	Development in time of assessment. "What is the ideal time for assessing students?"
Considering students while administering exams for reliability	6	Development regarding student factor and how does it influence reliability in administering exams.
Clear instructions for reliability in administering exams	5	Development in role of giving clear instructions in ensuring reliability in administering exams.
Readable test for increasing reliability in administering	1	Development in ways of increasing reliability in administering exams-understandable, readable tests.

According to the results, participants showed development in ensuring reliability during administering exams. First of all, it can be said that knowledge level of the participants about the suitability of the physical environment where the exam will take place has increased. Before the RRD sessions, the participants, who only paid attention to the seating of the students in order to prevent them from cheating again, stated that they would pay attention to the light, heat conditions of the classroom. Lecturers indicated that applying exams in a gloomy classroom or in an extremely hot/cold environment would prevent students from showing their competency, so reliability would decrease. After the RRD sessions, the participants reported that they would consider all these factors to increase reliability. The following excerpt illustrates this situation:

In a quiet environment, that is, in the classroom where the exam will be held, I ensure that the physical environment is suitable for the exam. This way I can ensure

credibility. *The classroom environment should be sufficiently bright. The classroom environment should not be exposed to outside noise. The classroom should be neither too hot nor too cold. All of these factors can affect students in revealing the knowledge they have. If all of these are at an optimal level, reliability will be ensured. (Comparison pre-semi-structured interview, P3).*

In addition, the participants stated that they applied exams without paying attention to the time of the assessment before the RRD sessions. It was found out that they learnt the influence of time of day on reliability, thanks to the RRD sessions. The participants stated that from now on, they will pay attention to the fact that the exam time should not be the first or the last lesson, not immediately after the meals, not immediately after long holidays or right after the physical education classes. Explanation of P3 sheds more light on this emerging situation:

We shouldn't assess after long vacations. These affect reliability. We have to pay attention to the environment. (Comparison pre semi-structured interview, P3).

The other thing which is totally new for me is the importance of when the exams should be taken. Almost we all know that morning hours are the best times for exams to be implemented, however, I did not know that after physical education class or after holiday the students should not have exams. Actually if you ask my real opinion, I think these times should be the most appropriate times because the students have rested minds and probably they are less anxious. Anyway, if I want to assess my objects, I will be more carefully about that detail as well. (Participant reflection report, P3).

4.4.8.2. Teachers' Developments in Factors Influencing Reliability in Assessment

Participants' self-perceived developments about the factors affecting reliability in assessment are gathered under this sub-theme. Based on the results, teachers' development regarding these results were found as assessment tool, assessor namely teachers, measurement error, and students, Table 99 presents these emerging codes:

Table 99.

Teachers' Developments in Factors Influencing Reliability in Assessment

Code	f	Code Description
Considering assessment tool while ensuring reliability in assessment	6	Development in participants' cognitions about how is reliability influenced by an assessment tool.
Considering assessor in ensuring reliability in assessment	6	Development in participants' cognitions about how is reliability influenced by assessors namely teachers
Measurement error in ensuring reliability	4	Development in increasing reliability by interfering with the measurement error.
Considering students in enhancing reliability	2	Development in participants' cognitions about how is reliability influenced by students who are being assessed.

According to the participants, quality of reliability is affected by the assessment tool, the assessor, the students, and the inclusion of measurement error in the assessment process. According to the participants, the assessment tool should be understandable, if the tool is a written exam, the writings should be legible with appropriate fonts. The participants also stated that they did not know that reliability in assessment was affected by the assessment tool, and that they would definitely take this into account in the future assessment practices to ensure reliability. The excerpt explains this situation:

*Yes. Actually, I was not enhancing reliability in my exams. I learned that **the assessment tool affects reliability**. We should **obtain similar results with the assessment tool, regardless of time and place, even regardless of the assessor, wherever we apply that assessment tool**. All students should understand the same thing from the assessment tool and act accordingly. An assessment tool that does not provide these negatively affects reliability (Comparison pre-semi-structured interview, P3).*

In addition, participants who learnt that teacher behavior in assessment also affects the quality of reliability developed their cognitions after the RRD sessions. The participants, who stated that they behaved more tensely, more seriously and nervously in the classroom environment during administering exams learnt that this behavior was wrong and they would not exhibit this behavior again. The participants, whose main aim was to increase reliability by behaving seriously throughout the exam, said that they violated reliability unconsciously, and they would not stress the students by acting calmly from then on and thus increase reliability. In addition, teachers stated that they should assess without integrating their feelings and personal thoughts into assessment, and that teachers who act biased in the assessment process reduce the quality of reliability.

Before starting the exam before the RRD sessions, I would begin the class with a tense and serious face. Giving instructions by shouting, acting nervously... While doing these things to ensure seriousness and reliability, it turns out that I was harming reliability without realizing it. I won't be doing this anymore. By creating a more peaceful environment, I will assess students without stressing them.
(Comparison pre-semi-structured interview, P4).

In addition, the instructors stated that the measurement error is the biggest factor reducing the reliability in the assessment process and that they would act carefully so that the reliability is not adversely affected by it.

4.4.9. Teachers' Developments in Measurement Error

In this theme, findings of the participants' self-perceived developments after RRD sessions in their cognitions related to measurement error were brought together. The theme consists of two sub-themes such as teachers' developments in knowledge of measurement error, teachers' development in how to prevent measurement error.

4.4.9.1. Teachers' Developments in Knowledge of Measurement Error

As a result of the analysis, it was revealed that the participants who gave superficial information about measurement error before the RRD sessions had detailed information about it after the RRD sessions. The codes that make up this sub-theme and detailed information about them are given in Table 100:

Table 100.

Teachers' Developments in Knowledge of Measurement Error

Code	f	Code Description
Increased knowledge in definition of measurement error	16	Development in the definition of measurement error. What does it account for?
Influence of measurement error	15	Development regarding influences of measurement errors on assessment, teaching and students' learning.
Increased knowledge in sources of measurement error	10	Development of participants' cognitions considering possible sources of measurement error, how it occurs.
Measurement error and reliability relation	5	Developments in the way of how measurement errors influence the reliability then validity.
Development of kinds of measurement error	4	Development in kinds of measurement error.

Participants, who learnt definition of measurement error in detail, stated that they had improved themselves on the subject and talked about possible effects of measurement error on the assessment process. Prior to RRD sessions, the participants could not make any connections among measurement error, reliability and validity. Participants who showed

development in their cognitions about measurement error after RRD sessions stated that it negatively affects reliability, it is not possible to ensure validity with unreliable results, and the decisions taken as a result of this process will not reflect the truth. According to the participants, in order to be able to make valid decisions during assessment process, first of all, it is necessary to obtain reliable results that do not contain measurement errors. The following excerpts summarize this situation:

*Measurement error seemed like an easy thing at first. It sounds like it can be compensated, but actually it can't. **The assessment must be error-free so that we can obtain reliable results. Not all reliable results are valid, but not all valid results are reliable. To ensure validity, we must first establish reliability.** We need to debug them. I want to add these. If we get good, valid data, we interpret it better. We get better results. We make education better. (Comparison pre-semi-structured interview, P4).*

*My explanation is correct but incomplete. I wasn't aware beforehand that reliability is so important in assessment. **The subject of measurement error is very important in assessment because the assessments we make must be error-free.** It should be error-free so that we can obtain reliable results. Because, thanks to reliable results, we can **make valid inferences and make valid decisions.** Thus, we can improve the quality of education. I can say that. (Comparison pre-semi-structured interview, P4).*

The participants, who could not address the relationship between measurement error and reliability before the RRD sessions, could easily establish this relationship after the RRD session they attended. They stated that when errors are involved in the assessment process, the data obtained will be erroneous and naturally they will not reflect the truth. Emphasizing the need to include as little measurement error as possible for reliable results, the participants stated that they would definitely prevent measurement errors in order to ensure reliability in their future applications. The following excerpt explicates the situation as follows:

*We discussed reliability and validity. In fact, in order **for the assessment results to be reliable, there should be no errors in the results during the assessment.** They are all interconnected. It really forces such a logic, but it is very enjoyable. **Reliable results are obtained through error-free assessments.** (Last pre-semi-structured interview, P3).*

4.3.9.2. Teachers' Developments in How to Prevent Measurement Error

In this sub-theme, there are research findings that refer to the development in the cognitions of the participants about preventing measurement error. They improved their cognition after RRD sessions regarding preventing measurement errors while ensuring reliability in assessment procedures. To illustrate, this emerging sub-theme includes codes such as using answer key for preventing measurement errors, avoiding measurement errors by foreseeing them and taking necessary precautions, and arranging assessment tool and physical environment of assessment to prevent measurement errors. Table 101 details the sub-theme:

Table 101.

Teachers' Developments in How to Prevent Measurement Error

Code	f	Code Description
Preventing measurement error through answer key and rater reliability	8	Development in ways of preventing inclusion of measurement error into assessment practices, Using an answer key and rater reliability
Preventing measurement error by foreseeing possible errors	4	Development considering ways of preventing inclusion of measurement errors into assessment: guessing possible errors and blocking them before they come out.
Preventing measurement error by being cautious in every stage	1	Development in preventing inclusion of measurement errors into assessment: teachers' responsibility, being alert in every stage of assessment
Preventing measurement error caused by physical environment	1	Development considering preventing inclusion of measurement errors into assessment practices: organizing, and preparing the physical context in which assessment practices will be held.
Preventing measurement error caused by assessment tool	1	Development considering preventing inclusion of measurement errors into assessment practices: assessment tool related errors.

When Table 101 is examined, it can be said that cognitions of the participants have developed in preventing the inclusion of errors in assessment. Participants who stated that they prevented the inclusion of errors in assessment only by preventing students from cheating could now prevent mistakes in many ways after RRD sessions. One of the participants' self-perceived developments regarding preventing inclusion of errors in assessment is to provide rater reliability. According to the participants, they can ensure consistency with intra-rater and inter-rater reliability and thus increase reliability by preventing the inclusion of measurement error. The explanation taken from one of the teachers' answers illustrates the situation:

*We can prepare a detailed answer key and be subject it to avoid confusion. Also, we can grade the exams again one week after studying the exams. **This is intra rater reliability.** Or we can ask another friend to rate it. **This is inter rater reliability.** If there is consistency, there is no problem. (Last post-semi-structured interview, P1).*

In addition, participants reported to have information about the types of errors and possible sources of error, so they can predict where and how the errors may originate and thus prevent the inclusion of errors in assessment by taking precautions. This finding can be clearly seen in one of the teachers' own sentences:

*We learned the **sources of error**, we discussed them. We learned and discussed what **kind of error can occur from which error source**. I know these. I am now aware of where and how an error can be included within the assessment process. We can **prevent the inclusion of error by preventing them and taking the necessary precautions**. (Comparison pre-semi-structured interview, P4)*

In addition to these, the participants stated after the RRD session they attended that assessment is a more comprehensive process than they thought and that the participants should have information about every stage. From the point of view of preventing the inclusion of errors in assessment, the participants indicated that they, as teachers, should act with knowledge and care in every step of assessment, so they can prevent mistakes. This result is exemplified in the quote:

*From question writing to test preparation, their administration, reading, evaluation and interpretation of the results, **an error may occur in all**. This should not happen. **We have to be careful**. Teachers should open their eyes. (Last pre-semi-structured interview, P1).*

As a result of the analysis, self-perceived developments of the participants considering issues apart from assessment were also found out. As a result of the analysis, these self-perceived development of the participants apart from assessment are related to Bloom's taxonomy, 21st century skills, vocabulary learning and being more planned in social life. When the results were examined, the participants stated that they mostly improved themselves about Bloom's taxonomy. They stated that they had heard about Bloom's taxonomy until the RRD sessions, but they did not have detailed information about it, but they learnt this taxonomy thanks to RRD sessions. The following statements demonstrate aforementioned development:

*We learned the levels in Bloom's taxonomy. I had heard about Bloom's taxonomy until the RRD sessions. **I am a graduate of the faculty of science and letters and I did not prepare for KPSS**. I have not had the opportunity to learn this taxonomy. But after the RRD session, **I learned about Bloom's taxonomy and its importance**. **I can consider it while planning my instructing and assessment**. (Last pre-semi-structured interview, P2)*

In addition, another self-perceived development that the participants showed apart from assessment is related to the learning of 21st century skills. The participants stated that they did not have any idea about what the main 21st century skills were and how they could be gained by the students until the RRD sessions. They claimed that they discovered 21st century skills thanks to reflective reading, and that they reinforced this subject through exchange of ideas in reflective discussions. The excerpt below clarifies the self-perceived development of 21st century skills of the teachers:

Lastly, thanks to RRD sessions, I have learnt 4Cs of 21st century skills such collaboration, cooperation, creativity and critical thinking. I also learnt that how these skills are employed in language teaching. I will search about them and learn more (Participant reflection report, P1).

In addition, the participants stated that they had a chance to learn new words while reading, so their vocabulary knowledge levels increased thanks to RRD as it is seen:

Also I have learnt new vocabularies. This reflective reading and discussion session improved my vocabulary. For example I did not know that plural form of curriculum is curricula. (Last pre-semi-structured interview, P4).

In this section, firstly, the research findings concerning teachers' current nature of cognitions about language assessment before the RRD sessions have been given. Secondly, the research findings revealing the strengths and weaknesses of teachers' language assessment knowledge and practices were shown, and finally, after ten RRD sessions, the findings regarding the development in the language assessment cognitions of the teachers were exhibited. In the following chapter, each research question will be discussed separately.

CHAPTER V

DISCUSSION

In this chapter, the data obtained are discussed with specific references to pertinent literature. Each research question is discussed separately. First, this chapter begins by providing a discussion in terms of teachers' current perception of assessment before reflective reading and discussion. Then it moves on to their strengths and weaknesses regarding assessment and evaluation. In the end, it focuses on the development of the teachers' cognitions and practices about assessment as a consequence of a series of reflective reading and discussion sessions.

5.1. Discussion for the Research Question 1

Research Question 1: What is the nature of EFL instructors' cognitions regarding language assessment?

The first research question in the current study is related to general perceptions of EFL instructors working at a state university. As a result of the analysis, we can say that teachers give importance to assessment and assess their students according to their objectives. Based on the research findings, it can be said that teachers perceive assessment as a procedure in which they utilise mostly written exams to explore students' understanding levels. In other words, teachers were found to associate assessment with administering tests to the students to shed light on whether they learnt the subject. In their answers, teachers constantly referred to points such as applying exams and determining the learning level of students. In addition, while the participants were talking about classroom assessment, their explanations included similar expressions, such as assessment as a process of determining students' understanding levels as a result of teaching. Additionally, especially novice teachers from the participant teachers perform assessment and evaluation practices in the same way as they were mostly evaluated when they were students. These results show the influence of teachers' previous learning experiences on teachers' perceptions and practices (Borg, 2003; Holt Reynold, 1992).

In addition to teachers' views about the definition of assessment, their perceptions were revealed as a result of the analysis regarding the ideal assessment time and technique, how an effective assessment should be, the purpose of assessment, assessment results, and the

country's and institution's assessment expectations from teachers. It can be said that all the results related to these issues are in the same direction. To elaborate further, teachers' definitions of assessment affect their perceptions and classroom practices about how ideal assessment should be, assessment purposes and effective assessment technique. On the basis of these findings, we can say that teachers, who associated assessment with applying written exams to students to determine their learning levels perceived written exams such as mid-term and quizzes as effective assessment. In addition, while all the participants talked about the purposes of assessment and evaluation, the teachers who spoke about checking students' understanding stated that written exams are an effective assessment method in connection with the definitions of perception assessment. Based on all these results, we can also say that teachers adopt traditional assessment in their practices. This finding is in line with what Volante & Fazio (2007) figured out in their studies. To illustrate, their study revealed that teachers utilized testing and assessment tools for traditional summative purposes. There can be many reasons behind the teachers' opinion that the written exam technique is ideal for effective assessment. These beliefs may have resulted from such factors as teachers' habits, practicality, affordability, and institutional requirements.

As for the ideal assessment time, teachers' opinions were revealed as being in line with the summative assessment. To go into further detail, nearly all teachers assess their students at the end of a period spent teaching and evaluate their students to make decisions about their success. When teachers talk about the ideal time for assessment and evaluation, the fact that they indicate the end of the semester or a few units reveals that they have adopted a result-oriented assessment as Volante & Fazio (2007) figured out. Based on this, we can suggest that this situation might be shaped both by teachers' adoption of the traditional assessment and by the assessment experiences they had during their student years (Borg, 2003; Holt Reynold, 1992). Additionally, during the investigation, one of the teachers stated that she had a process-oriented assessment approach. Later in the study, it became clear that s/he did not know enough about formative assessment. With this result, the importance of teachers' assessment knowledge has emerged as Jin (2010) pointed out that teachers' should be knowledgeable about assessment to apply it appropriately. Also this finding confirms the fact that a teacher with sufficient assessment knowledge and skills is also required to appropriately implement this knowledge into his/her assessment-related practices. Although teachers are predisposed to formative assessment, their incomplete knowledge prevent them apply what they intend to do.

Therefore we can say that this disposition should be supported by teacher assessment knowledge.

All teachers were found to give importance to assessment and evaluation results. Before the RRD sessions, the teachers used the results only to determine the students' learning status for making decisions about student achievement and how effective they were to students as teachers. The results show that the teachers who stated that they would organise their teaching have formative assessment intentions in assessment and evaluation. However, in the later stages of the research, it was found that the teachers did not take any action to interpret the assessment results obtained from the students; in short, they were found to have an idea by looking roughly at the results. This situation shows that teachers' assessment and evaluation knowledge levels should be at the desired level (Jin,2010) to perform assessment suitably.

Another reason why teachers could not use assessment results having formative purposes can be cited as the workload of the instructors. Teachers with a heavy course load spend most of their time and effort on teaching English and cannot spare much time for assessment. Although they are aware of the assessment's importance, teachers prioritise teaching. In addition to these, although teachers intend to implement formative assessment, they cannot fully achieve it; the expectations of the school and the country's education system from teachers concerning assessment emerged as another factor as a result of the analysis. To illustrate, one of the teachers stated that s/he intended to apply the formative assessment and use alternative assessment techniques in the first years of their career. Still, the school required students to be assessed through written exams only during midterm finals, which caused him not to be able to put his plans into practice. As it can be seen, although the teachers have different plans, institutional expectations from the teachers concerning the assessment cause them to perform in a way that is different from their intentions and goals. In fact, the participant stated that over time, his expectations of the school began to seem reasonable to him, and he believed that they were correct and acted in line with the school's expectations in his practices. This emerging results is parallel with what Burns(1996) concluded that organizational requirements and institutional context influences how the teachers plan and act accordingly. In short, this situation shows us that the assessment policy of the school and the country is effective on the assessment perception levels of the teachers and enables them to be reshaped over time.

5.2. Discussion for the Research Question 2

Research Question 2: What are participants' strengths and weaknesses considering language assessment as they perceive it?

This section of the chapter includes a discussion of the findings referring to the strengths and weaknesses of EFL instructors in the assessment process. When instructors were questioned about their assessment strengths and weaknesses, they responded more frequently to the questions regarding their weaknesses. In other words, based on the data, we may conclude that the number of instructors' weaknesses was more than their strengths. Even some of their self-perceived strengths were discovered to be insufficient or lacking. Therefore, it may be deduced that instructors had more shortcomings in assessment and evaluation than strengths (Tzagari and Voght, 2017; Sultana, 2009; Volante and Fazio, 2007). Similarly, Wach's (2012) study, which attempted to investigate university EFL teachers' self-perceived strengths and weaknesses in assessment, identified several weaknesses of the teachers regarding assessment, even though their strengths regarding this issue were found to be limited in number.

In this study, teachers' assessment strengths emerged related to assessment results, assessment techniques, developing and administering exams, basic introductory terms and issues related to assessment, and ethical considerations in assessment. As for their weaknesses, they showed themselves in various topics such as test construction and item writing, alternative assessment, basic introductory terms, assessment results, assessment techniques, grading and scoring, ethical considerations, and language assessment principles.

When asked about their strengths in assessment, the instructors' one of main concerns was found as reliability. The fact that teachers mostly state that they are equal and objective, that they administer exams by preventing students from cheating which in turn will increase the reliability, and that they give grades equally and objectively to all students, which is again believed to increase the reliability led to the emergence of this finding. While the teachers were addressing their practices, they did not explicitly indicate that they implement them to ensure and increase reliability. However, the characteristics they highlighted as their strengths and the practices they performed led to the emergence of this result. These findings suggest that teachers prioritise reliability in the language assessment procedure. Based on emerging findings, the data revealed that instructors used the term objectivity to characterise their strengths in increasing reliability in their activities. Objectivity may be described as a goal of

assessment characterised by freedom from subjective effects in general and the capacity to view the situation correctly without the influence of emotion, prejudice, or bias (Van der Vleuten, Norman, & De Graaff (1991). When teachers' own words are considered, we can say that their priorities are given only to prevent cheating, which may result from their early experiences as learners. To illustrate, teachers who were tested and assessed under strict and serious circumstances created to hinder students' cheating led them to conceptualise reliability like this.

Teachers also mentioned their strengths in terms of validity. Throughout the interviews, teachers only touched upon content validity when they were asked about their strengths regarding assessment. They were found to take only content validity while preparing tests and choosing or adapting assessment techniques or tools appropriate for their objectives. The findings are consistent with Shim's (2009) findings, which revealed that most English instructors indicated they paid attention to content validity and assessed their students based on what was taught in the lesson and the activities and methods used in the classroom.

One of the emerging strengths of the participants regarding assessment is their consideration of assessment as a process. Rather than focusing on the product, teachers believed that assessment should be conducted by emphasising the process by implementing the alternative assessment. The interviews also found that few instructors viewing language assessment as a process utilised alternative assessment strategy such as portfolio assessment, which allows teachers to track students' progress over time. However, they were found to have limited knowledge of alternative assessment techniques. Xie and Tan (2019) reached a similar result when they discovered that although teachers are aware of the need to implement formative assessment, they lack a complete understanding of it.

5.3. Discussion for the Research Question 3

Research Question 3: Do reflective reading and discussion sessions lead to any development in EFL teachers' cognitions concerned with language assessment, as they perceive? If so, what are the changes?

After participating in a series of reflective reading and discussion sessions, the content of which was determined according to the self-perceived weaknesses of the participants, teachers participated in two different semi-structured interviews. Then, teachers were observed in their classrooms and were included in follow-up semi-structured interviews. The

data analysis showed that participants developed their cognitions considering assessment. To illustrate, emerging teachers' self-perceived developments were revealed related to assessment techniques, test construction and writing test items, assessment results, ethical considerations, administering exams, and assessment principles.

The first and most important conclusion drawn from the data gathered following RRD sessions reveals that the instructors' concept of assessment developed with the help of reflective reading and discussion. They began to feel more confident, competent and informed about language assessment.

Next, teachers made progress in choosing an assessment technique or tool suitable for their goals. Before the RRD sessions, it was revealed that when deciding on an assessment technique or tool they would use in their lessons, they considered possible experiences they might have while applying it, its practicality and content validity. Namely, it was discovered that instructors prioritise practicality and experience more, which is backed up by Yıldırım's (2012) research. To illustrate, when participants chose assessment techniques for their objectives, they preferred techniques that students were acquainted with since they did not want to have difficulty administering them. Davison (2004) discovered similar findings and noted that teachers might have negative experiences when using assessment methods owing to their students. As a result, they tend to select and apply assessment methods that students are comfortable with. After the RRD sessions, it was revealed that when the teachers choose an assessment technique for their practices, they will consider both their objectives, target students and assessment principles.

Secondly, participants reported an increase in their knowledge considering the purposes of the assessment. Before the RRD sessions, participants assessed their students to check their understanding and collect data to determine their success. Also, they reported evaluating to self-criticise themselves about their teaching efficacy. The findings of Herrera and Macias (2015) are directly in accordance with this finding. They identified that language assessment could be employed to examine students' learning, decide who passes and fail, and self-assess teaching. After the RRD sessions, development in participants' cognitions regarding this issue emerged, and the participants showed a tendency to assess for learning. This may be because teachers are inclined to employ and profit from assessment to enhance their teaching and student learning. This is significant because teachers will utilize assessment to enhance their teaching techniques and students' learning, rather than to score students' performance (Giraldo 2018). Although participants' assessment aims were directly related to merely assessing

performance or making a pass-fail decision before RRD sessions, instructors confirmed that their assessment aims shifted toward evaluation for improvement and learning. The results of Chappuis and Stiggins (2016) about how instructors should perceive and implement assessment corroborate this finding. As Brown (2004) said, this emerging result is in accordance with the claim that teachers should know about assessment purposes such as proficiency, placement, achievement and diagnostics.

Teachers should know various assessment techniques and their strengths and shortcomings when selecting assessment methods (AFT et al., 1990). Regarding this claim, it can be said that teachers' knowledge of assessment techniques has considerably increased after RRD sessions. To illustrate, teachers' thoughts regarding assessing students in language classrooms shifted from traditional methods to alternative ones. At first, the participants only knew about traditional methods of assessment. At the end of the RRD sessions, however, most of them said they were in favour of using process and performance-based assessment techniques in language classrooms, which are seen as an important priority for effective teaching and learning (Stiggins, 1999).

As for teachers' cognition development related to administering exams, it can be said that there is a visible increase in teachers' knowledge concerning administering exams, scoring and grading them. Before the RRD sessions, participants used to consider preventing cheating while administering exams. However, participants indicated their development in areas such as reliability, the physical environment of the classroom, giving instructions, and being effective proctors. When participants were observed while administering exams, they displayed overlapping behaviours in a well-established procedure. This development is in the same direction as what MAC (2013) recommends. To go into detail, Teachers began enhancing the security of their exams and avoiding any issue that may invalidate assessment findings and influence students' actual skill levels. In addition, teachers became acquainted with reliability and its impacts on assessment, so they take precautions such as organising seating, providing clear instructions to students, establishing an exam environment with appropriate temperature and lighting, and minimising measurement errors.

CHAPTER VI

CONCLUSION

6.1. Introduction

This chapter reviews the goals of the study, clearly responds to the research question(s), and summarises the main findings. The implications and limitations of the study are also covered in this chapter, along with suggestions for further inquiry.

6.2. Overview of the Study and Implications

The primary purpose of this study was to focus on the assessment-related cognition development of English language instructors as a result of reflective reading and discussion. Furthermore, this research investigated teachers' cognitions concerning assessment and found out participant teachers' strengths and weaknesses regarding assessment. In line with these goals, the following research questions are formulated;

1. What is the nature of EFL instructors' cognitions regarding language assessment?
2. What are participants' strengths and weaknesses considering language assessment, as they perceive it?
3. Do reflective reading and discussion sessions lead to any development in EFL teachers' cognitions concerned with language assessment, as they perceive? If so, what are the changes?

Various data-collecting tools were employed to obtain answers to these formulated research questions. These instruments included pre- and post-classroom observations, individual semi-structured interviews (pre and post), KWL charts, and reflection reports.

In light of these aims, this study was divided into three phases: before, during, and after RRD sessions. The research began with the pre-RRD sessions phase to shed light on participants' current cognitions concerning assessment. Namely, teachers' initial assessment perceptions were explored before participating in any RRD session. Following this, the second stage of the research, during- RRD sessions, consists of collecting data regarding teachers' self-perceived strengths and weaknesses in assessment, planning and designing an RRD session program then putting it into practice. Four different data collection tools were

employed in this phase. The first was individual pre-semi-structured interviews aiming at exploring participant instructors' strengths and weaknesses concerning assessment. The KWL chart was another tool utilized during RRD sessions to detect teachers' existing knowledge, express their expectations from both reading and the RRD sessions, and enhance awareness of their learning gains at the end of the sessions. Also, reflection reports were employed by participants and the researcher. This phase, on the other hand, intended to understand in which areas of assessment teachers perceived themselves as more knowledgeable and in which areas of assessment they perceived themselves as having incomplete or lack of knowledge. As a result, this study phase can be seen as a means of needs analysis. The last phase of the research intends to reveal the development in teachers' cognitions regarding assessment as a result of RRD sessions. Also, to what extent these self-perceived developments are reflected in participants' actual assessment practices has been aimed to be identified. As a consequence of this comprehensive research, the following list summarises the key findings:

Before the RRD sessions, each participant attended a pre-semi-structured interview to reveal the nature of the teachers' assessment perceptions. Afterward, the teachers were observed in their classrooms, taking into account their statements during the interviews. Then, each teacher was again included in a follow-up interview. As a result of the data analysis obtained after this procedure determined the teachers' definitions of assessment, their views on how an effective assessment should be, and their thoughts on classroom assessment. In addition, the statements of the participating teachers revealed their perceptions about the purposes of the assessment and the appropriate time for the assessment. Finally, teachers' self-perceived reports about the perception of assessment in the first years of their teaching career and how it developed over time emerged among the analysis results. To go into detail, teachers see assessment as applying a written exam to students to determine whether students have learned the subjects at the end of a certain process spent with teaching. When the teachers' answers were examined, it was revealed that both assessment and classroom assessment were perceived similarly and that teachers' assessment definitions did not go beyond administering written exams to students. Teachers see this practice as an effective assessment. These results suggest that teachers generally perceive and practice assessment based on traditional assessment.

Looking at the schemata of the teachers before the RRD session in assessment and evaluation techniques, it was revealed that they mostly knew traditional assessment techniques and included them in their classroom practices. Teachers who perceive alternative

assessment only as an assessment technique have also shown that they lack knowledge about such assessment methods. In addition, it was found that teachers mostly use assessment techniques to determine students' learning status. Almost all teachers indicated that the assessment technique they used in their practice was the written exam, and when asked, they stated that they preferred that technique because it was practical for both students and teachers. As a result, the research findings referring to the teachers' perception of general and assessment techniques and the use of assessment techniques were found in the same direction.

It has been discovered as a result of the analysis that teachers consider the assessment results obtained through assessment and evaluation practices as important and use the data to improve their teaching and students' learning. In other words, it was revealed that participants' purposes of using assessment results were limited only to improving students' learning, revealing their strengths and weaknesses as teachers, and making arrangements in their teaching for more effective teaching practices.

In addition, the assessment expectation of the institution teachers work in, and the country's education policy effectively shaped the participants' assessment perceptions. The teachers who stated that they aimed to adopt a more process-oriented assessment at the beginning of their profession and to use alternative assessment methods by taking into account individual differences said that the school's assessment expectations from them and the education policy of the country made them apply mostly written exams at the end of the semester or year to decide on the success of the student. Over time, these practices became plausible for them.

In other words, the assessment expectations of the institution where the teachers work and the country's education policy changed their perceptions about the definition of assessment, the purpose of assessment, the ideal assessment technique and the ideal assessment time, and reshaped them in time.

In addition, before the RRD sessions, it was aimed to reveal the strengths and weaknesses of the instructors related to assessment and evaluation. For this purpose, semi-structured interviews were arranged with each teacher, and then the teachers were observed in their classrooms. Pre-created observation schemes were used during the observation, and the researcher kept field notes. Based on the results obtained from these two procedures, a follow-up semi-structured interview was arranged, and this interview was applied to each teacher separately. Teachers' strengths and weaknesses related to assessment were explored from the data obtained.

The first strength of teachers to be revealed is related to their use of assessment results. Teachers use the results to determine students' learning status and, accordingly, to make decisions about student achievement. In addition, it was determined that the teachers used the assessment results to evaluate their teaching. In other words, it was found that the perceptions and practices of teachers as a result of assessment overlapped with each other. Although the teachers were not observed to use the assessment results during the observation, in the follow-up interviews, the teachers confirmed that they used the results only to evaluate their teaching and to decide on their success by determining the student's learning level.

Teachers' strengths related to assessment techniques were classified, such as their knowledge of assessment techniques and their implementations, the factors that teachers consider when choosing a technique, and the purposes of teachers' use of assessment techniques. Teachers were explored to generally know and apply written exams. Teachers primarily consider practicality and students' familiarity when selecting assessment techniques for their practice. In addition, teachers use assessment techniques to determine whether students have learned the subjects. This result shows that the teachers' perceptions of the purpose of assessment and the pursuit of using the assessment technique are in the same direction. In addition, a clear pattern emerged in teachers' general assessment perceptions. Teachers tend to focus on students' perceived definitions and assessment purposes, their use of assessment techniques and the derived assessment data.

Teachers' strengths related to test construction and administering exams were also determined as a result of the analysis. Teachers only consider content validity while preparing tests. It was also found that they increased the reliability of their exams by preventing students from cheating during applying the exams they prepared.

Teachers' strengths on ethical issues in the assessment are limited to grading only. Teachers ensure ethical behavior by treating each student equally and objectively while grading student performances or written papers. The teachers' strengths in grading and scoring are similar to the findings concerning ethical issues. In other words, the teachers only mentioned being objective as their strength in scoring.

In the study, teachers' weaknesses related to assessment and evaluation were also revealed. For this purpose, semi-structured interviews, classroom observations and follow-up interviews were used as data collection tools. In addition, the results obtained are one of the criteria used to determine the content of RRD sessions. We can say that this application might be seen as a need analysis related to the assessment and evaluation of teachers.

The first self-perceived weakness to be revealed is related to teachers' test construction and test item writing knowledge. As a result of the analysis, it was determined that the teachers have incomplete knowledge regarding what to consider when constructing a test and that they lack knowledge of various test items and their writing. In addition, teachers stated that they had weaknesses in item analysis.

Secondly, teachers' perceived weaknesses related to alternative assessment were identified. These weaknesses are related to teachers' knowledge of alternative assessment types, performance assessment, knowledge of rubric design, and implementation of some specific alternative assessment methods. As a result of the analysis, it was determined that the teachers misunderstood the performance assessment and applied it incorrectly. In addition, it has emerged that the teachers have little knowledge of alternative assessment methods. Therefore, they either do not use alternative assessment methods in their practices or make incomplete applications when using them.

In addition, it has been revealed that teachers are confused about some basic terms related to assessment and evaluation. It has been observed that teachers, who limit the purpose of assessment only to identifying the learning levels of the students, also have weaknesses in assessment types. It shows that teachers who state that formative assessment, which is one of the assessment types, is an assessment technique need to develop knowledge about the subject.

As for the weaknesses of the teachers related to the assessment results, the teachers stated that they used the assessment results when interpreting them, as seen in the previously revealed findings, not only to determine the learning level of the students but also to evaluate the quality of their teaching.

It has also been revealed that the teachers have weaknesses in the criteria to be considered when choosing an assessment technique, in the knowledge of the kinds of assessment techniques and their implementation, and in the purpose of using these techniques. It has been observed that teachers' knowledge of assessment techniques is limited only to traditional assessment techniques. In addition, teachers used assessment techniques only to check students' understanding levels and did not mention other possible purposes. Finally, it was revealed that teachers consider practicality and content validity when choosing or developing the assessment technique.

As a result of the study, it was revealed that the teachers had a superficial knowledge of a few assessment principles while they did not know some of them at all. Teachers, who explained reliability only by preventing cheating and being objective, only mentioned content validity when talking about validity. In addition, it was revealed that the teachers did not know about washback and had some weaknesses regarding assessment principles.

Finally, it was revealed that teachers perceived grading and scoring similarly. In addition, while teachers were talking about grading, it was revealed that their explanations were limited to being objective and equal while grading students. Last but not least, research results show that teachers also have incomplete knowledge about ethical issues in assessment. Teachers, who only mentioned objectivity in grading when talking about ethics, did not mention it at all when talking about other stages of assessment. These results show that teachers also have weaknesses in grading, scoring and ethics in assessment.

After the teachers' weaknesses were discovered and their opinions were taken, a series of sessions were planned for reflective reading and discussion. The content and order of these sessions were determined according to the results of the data obtained and in line with the views of the participant, the researcher and his supervisor. As a result, 10 RRD sessions were planned, and each session was conducted an average of 2 weeks apart.

Different data collection instruments were employed to reveal the development of the teachers' cognitions after the RRD sessions they attended. These are KWL charts, reflection reports, pre and follow-up semi-structured interviews and classroom observation. As a result of the analysis, it was revealed that the participants' assessment-related cognitions improved, and their stated weaknesses turned into their strengths. Teachers' perceived developments were found to be related to knowledge and implementation assessment techniques, including alternatives, preparing tests and writing test items, using, interpreting and evaluating assessment results, administering exams, ethical issues in assessment, increasing the effectiveness of assessment and evaluation, assessment types, assessment purposes, assessment principles, grading and scoring.

Participants indicated in their statements that assessment is more comprehensive and important than they thought. It was also found that the teachers who reported realizing that their assessment knowledge was insufficient before the RRD sessions largely eliminated their weaknesses in the subject thanks to the RRD sessions.

Although the progress claimed by the participants could not be observed in the observations, the development of the teachers after the RRD sessions was confirmed by the analysis of the data obtained from the follow-up interviews after the observation.

6.3. Implications of the Study

Borg (2019) contends that research on language teacher cognition can only be successful if it yields implications for language teacher education. How individuals learn to teach languages is a fundamental subject that remains at the center of ongoing debate (Borg, 2006; Li, 2017; Öztürk, 2015). It is possible that understanding this process through research on teachers' cognition might assist in developing teacher education programs. Kubanyiova (2015) proposed combining teacher cognition, teacher learning, and teacher education research. Since the current study focused on these constructs, its findings may provide pre-service and in-service teachers with confirmation of their assessment comprehension and knowledge. According to Johnson (2009), this sort of study on the cognition of language teachers might be utilized for both pre-service teacher education and continuous professional teacher development. Also, it is worth noting that teachers' educational perceptions are often informed by their general personal perceptions. Therefore, efforts should be made to assist them in establishing positive perspectives on assessment, extending their horizons and encouraging them to adopt suitable assessment procedures in their teaching.

In the present study, all the teachers considered language assessment equal to language testing and measurement, and they did not differentiate the terms testing, assessment, measurement and evaluation. They also tended to think language assessment refers to summative rather than formative assessment. To broaden their perspectives about assessment and to make them comprehend other functions of assessment, such as how assessment contributes to learning/teaching (Jin, 2010; Klinger, 2016), the content of the courses given in pre-service education should be revised, or appropriate in-service teacher education should be offered to in-service teachers. It is because teacher education shapes teachers' beliefs and perceptions about assessment according to what they learn, which affects their future assessment performance in their classes (Yan, Zhang & Fan, 2018).

Firstly, the current study indicated that teachers at a state preparatory program preferred traditional assessment strategies such as written exams and that existing assessment processes were teacher-oriented. This may have resulted from curricular requirements and current classroom settings or the fact that instructors view traditional assessment techniques as more

trustworthy and practical than other types. More formative assessment types, such as portfolios and self-assessments, might be suggested for assessing students' language competencies, informing teachers that this type of assessment is more likely to increase students' learning gains while learning a foreign language.

In the present study, it was discovered that instructors behaved according to institutional expectations concerning assessment, particularly in selecting assessment techniques, although expressing that they did not appreciate these institutional requirements regarding assessment. This result can be interpreted as discomfort felt by the teachers during their assessment practices as Berry, Sheehan, & Munro (2019) argue that this sense of discomfort is not caused by a lack of assessment knowledge as the results of studies would suggest (Fulcher, 2012; Vogt & Tsagari, 2014). Rather it is, we would contend, based on awareness of the limitations of assessments they were required to use since the teachers felt that there were other, perhaps better, ways of assessing their students (Vogt & Tsagari, 2014).

Next, administrators should give teachers enough flexibility to take the initiative concerning their assessment practices. Besides, teachers should be given opportunities to select the most appropriate assessment tool or technique related to their objectives. Therefore, teachers might believe they can upgrade the quality of learning and teaching; as Bandura (2001) states, people will not be motivated to take action if they do not feel capable of leading to the desired improvements.

Lastly, from the students' viewpoint, it is observed that having a language assessment literate teacher has certain advantages. First, such teachers can interpret the assessment results, identify students' needs and take actions to meet those needs, which will be resulted in improvement not only in teaching but also in the learning process. Furthermore, a teacher who practices self-reflection will be receptive to criticism, paving the way for a more democratic atmosphere in the classroom where students may express themselves freely. Moreover, assessment-literate instructors tend to inspire their students by incorporating them into assessment procedures, which benefits student motivation.

Moreover, it was evident in the interviews that the EFL instructors did not effectively use self-peer-portfolio assessment and gave more importance to summative assessment. Instructors must be informed that formative assessment significantly improves students' language achievement. They should equip themselves with these current and alternative assessment methods.

Training and professional development programs could be designed based on theory and practice related to language teachers' needs. They could be supported and encouraged to attend conferences and professional development programs on language testing and assessment. For the practical implication, the heads of the preparatory programs or the school principals could administer this scale to the language teachers working in their institutions. Based on the findings derived from the scale, the heads or principals could detect the weaknesses and strengths of these participant teachers and also could determine these teachers' needs regarding each skill. Based on the strengths, weaknesses, and needs of the teachers, professional development programs could be determined, and the teachers could be encouraged to attend conferences, and regular meetings could be held to exchange information. Thus, the language assessment knowledge of EFL teachers could be increased by being context-specific and considering institutional factors.

Pre-service education appears to have some limitations especially from the practical standpoint in language testing and assessment, and pre-service teachers are not equipped with the necessary knowledge in pre-service education related to language assessment. Thus, the content of the course in pre-service education might be considered to be revised. Moreover, one course cannot be sufficient for such a broad topic to be covered in just one academic term. There should be more than one course related to language assessment, and more practical hands-on practice can be incorporated into these courses in pre-service education. Also, English teachers who have graduated from other language departments apart from the ELT department should be given comprehensive in-service training based on assessment and evaluation in language teaching for the same purpose. These trainings should include a knowledge basis and actual implementation to facilitate assessment literacy (Deluca & Klinger, 2010). They should also be encouraged to reflect on their own assessment experiences and how these have influenced their assessment practice. Teachers should be encouraged to share their assessment knowledge and practices with colleagues. The findings indicate that these teachers favored adopting reflective practice through which they could enhance their learning by reflecting collaboratively and using reflective practice tools in the endeavour. As Rodgers (2002) and Zeichner and Liston (1996) posit, reflection should be done in interaction with others. These teachers seemingly had found rare chances to reflect collaboratively. They were appreciative of how their togetherness worked effectively with reflective practice, which suggests that when reflection is the case, collaboration is a needed. This can remind us of Dewey's (1933) idea of reflective activity – as opposed to routine

activity – that eliminates repetitive behavior and engages in enriching inquiry of practice by following a course of reflecting in, on and for action (Schön, 1983, 1987). Teachers should engage in collaborative reflection by consciously following a cycle of reflection in, on and for action (Schön, 1983, 1987). Teachers should stay open-minded, wholehearted and responsible (Dewey, 1933) for their development as a professional

6.4. Limitations of Study

This study aims to examine EFL instructors' cognitions and practices concerning assessment. The study is restricted by space and is therefore confined to EFL teachers at a school of foreign languages at Bingöl University. The study is likewise restricted by time and only covers the fall semester of the 2019-2020 academic year. Boundedness is a phenomenon that is explained by the exploratory nature of qualitative case study research, which is defined as the "in-depth description and analysis of a bounded system," and the particularistic characteristic of case studies, which means that "case studies focus on a particular situation, event, program, or phenomenon" (Merriam, 2009, p. 43). Next, the sample size of the current study, which consists of four English language instructors, might be another limitation. However, it should be noted that transferability is more important in qualitative studies than generalizability. The transferability of qualitative studies depends on the similarity of the context in which the research is conducted to the contexts of the studies (Mackey & Gass, 2005). Due to the familiarity of the contexts, similar findings can be obtained in other groups of EFL teachers teaching at some other schools of foreign languages in Türkiye. Furthermore, thick descriptions in presenting the study setting, participants, the researcher's role, and the data collecting procedure might be a strength in identifying context similarity.

6.5. Suggestions for Further Studies

Keeping the limitations in mind, this part of the research provides recommendations for further studies. First, this study contributes to the teacher cognition literature by studying a rare topic: teacher cognitions associated with assessment development. This study describes EFL instructors' assessment-related cognitions and practices, identifies their strengths and limitations in assessment, and demonstrates, if any, development in teachers' assessment-related cognitions as a result of reflective reading and discussion. Still, additional components might be added to further research. For instance, teachers' cognitions and practices in

addressing assessment challenges, the link between their cognitions and practices about the issue, or instructors' emotional tendencies in assessing students might be researched further.

Secondly, the present study attempted to explore EFL instructors' assessment cognitions and the development of these cognitions along with practices resulting from a series of RRD sessions. As a longitudinal inquiry, the investigation covered only an academic semester. This research might expand by carrying out a continuum or follow-up study with the same participants. This way, validity and reliability concerns could be strengthened, and the sustainability of teachers' self-perceived cognition development could be examined from a more in-depth and long-term viewpoint. In addition, by means of a follow up study, whether or not the participants actually put their newly developed theories as a consequence of this research study into practice may be observed.

As a result of this study, it should be noted that the researcher is cautious about making broad generalizations but instead gives a modest recommendation for all people involved in the assessment process. This study is believed to shed light on or provide guidance for other practices. To illustrate, future studies may carry out similar teacher cognition-focused studies to determine its feasibility in various contexts, such as high school, secondary school, and private school settings. To enhance the generalizability of the current findings, additional research might concentrate on re-production research by replicating the study design and methods in various institutional contexts and EFL environments.

Next, materializing the research suggestions provided above, answers to two key questions: "(1) What strategies do language instructors use to create relevant learning environments for their students? (2) How can teacher education and continued professional development help language instructors learn this way?" (Kubanyiova and Feryok, 2015, p. 435) might be found out. Further study might help bridge the gap between teacher cognition, teacher learning, classroom practice, and student learning. This research might also be useful in developing guidelines to promote assessment procedures in EFL classes and assessment training for major stakeholders.

In addition, it is suggested that research be carried out on the cognitions and practices of English language instructors in connection to various facets of instruction, such as corrective feedback, learner uptake, or classroom management. The number of elements of teaching that might be examined can be increased.

Finally, the participating teachers stated that after the RRD sessions, a workshop should be organized during which they could work with each other and practice what they learned on subjects such as multiple-choice test item writing, test construction, preparing an analytical or holistic rubric, and item analysis. Future research can add arrangements to their research procedures that will create an opportunity for teachers to work together and apply what they have learned.

6.6. What the Researcher has Learnt as a Result of Reflection and Discussion

In this part, development in the cognition of the researcher, who participated in the RRD sessions with the participants, was displayed after the sessions. After each RRD session, I, as a researcher, prepared a reflection report about the contribution of the sessions to me, just like the participants. As a result, I classified my cognition development regarding language assessment into seven broad areas such as developments in assessment principles, developments in preparing tests, developments in basic introductory topics regarding assessment, assessment result-related developments, additional assessment-related developments, developments in performance assessment and rubric, and developments in scoring and grading.

After the RRD session, I realised that I developed my cognition about the role of the teachers in ensuring reliability in administering exams. As a teacher, I used to be nervous and harsh to seem serious, just as the participants mentioned, while applying the exams in my assessment practices. In the RRD session I attended, I learned that this behaviour negatively affected reliability. From now on, I will be calmer while applying the exams and contribute to increasing reliability by creating a peaceful environment without stressing the students. In addition, another development related to reliability was about the physical context in which the exams were applied, affecting the reliability quality. After RRD sessions, I will be careful about the fact that the classroom environment is suitable in terms of temperature, light and sound. In addition, I showed development in my cognition on rater reliability. I did not know about inter-rater and intra-rater reliability and naturally did not apply them in my practices; I learned about these reliability types after RRD sessions. If consistency is ensured, reliability will also be provided. The other development was about washback. I learnt what washback is, how it is enhanced, and how it affects the assessment process, teaching and learning. Now, I know the importance of washback and how to increase it. I will ensure it at all stages of the assessment whenever possible.

Considering test construction, I learned what test formats are. Before the RRD sessions, I did not know the subjective or objective tests and whether to add the supplied test item or selected item test items to these test types. As a result of the sessions, I learned them. Next, while preparing the test, I learned how to order the test items according to their types and degree of difficulty. Additionally, I was able to increase my knowledge regarding test item writing. The most prominent development was that the words used in the item writing are determinants of the learning habits of the students. I have learned that using wh-questions such as what and when in exam questions prompted students to memorize, the questions including, for instance, why, requiring them to use high-level thinking skills. In addition, I showed significant improvement in multiple-choice test item writing. When I decided to use a multiple choice test item in my exams before the RRD sessions, I either downloaded it from the internet or prepared it quickly. I learned the rules such as having a diagnostic function of the options.

Moreover, I improved myself regarding alternative assessment. For instance, I realized through reading that I had misinformation about classroom observation. Thanks to the session I attended, I corrected my mistakes and will now apply observation, which is an alternative assessment method, in my assessment practices in a more knowledgeable and systematic way.

When comes to basic issues in the assessment, the definition of measurement was new to me. I learned that measurement is about quantity. In addition, there was a significant improvement in my perception of assessment after the RRD sessions. Considering all the stages of the assessment, I realized that the assessment was actually more comprehensive than I thought and that it was a subject that teachers should have knowledge of.

In addition, I did not perform any statistical calculations before the RRD sessions while interpreting the assessment results. After the sessions, I learned to calculate the item analysis, item difficulty level and item discrimination level. After that, I will use the data obtained in the light of statistical processes while interpreting the results.

Moreover, before the RRD sessions, I used to associate ethical issues with being fair and objective in assessment. After the RRD sessions, I learned that ethical issues are a very broad concept. I learned how to ensure ethics in all areas such as determining the objectives, preparing and applying the exams, scoring, grading, and interpreting assessment results. In addition, I realized the relationship between content validity and ethics.

Last but not the least, I learned new information about scoring and grading. Before the RRD sessions, the researcher used to act randomly while assigning scores to the questions in any

exam I would administer. After the RRD sessions, I learned what to consider while scoring and will materialize it effectively and accurately. When it comes to grading, I learned that I could assess students in several ways just like triangulation so that I would provide triangulation of grading and enhance reliability.



REFERENCES

- Abdal-Haqq, I. (1998). Constructivism in Teacher Education: Considerations for Those Who Would Link Practice to Theory. ERIC Digest.
- Abdullah-Sani, A. S. Z. (2000). An analysis of the development of teacher belief constructs during teaching practice and in the novice year of teaching: a case study of English language teachers in the Malaysian context (Doctoral dissertation). Retrieved from <http://wrap.warwick.ac.uk/36424/>
- Abtahi, Y. (2017). The ‘More Knowledgeable Other’: A necessity in the Zone of Proximal Development? For the Learning of Mathematics, 37(1), 35-39.
- Akbari, R. (2012). Validity in language testing. The Cambridge guide to second language assessment, 30-36.
- Akinoğlu, O. (2013). Yapılandırmacı öğrenme ve coğrafya öğretimi. Marmara Coğrafya Dergisi, (10).
- Alkharusi, H. (2011). Development and datametric properties of a scale measuring students' perceptions of the classroom assessment environment. International Journal of Instruction, 4, 105-120.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. Retrieved from <http://buros.org/standardsteacher-competenceeducational-assessment-students>.
- Andrews, S. (2007). Teacher language awareness. Cambridge: Cambridge University Press.
- Atkins, L. and Wallace, S. (2012). Qualitative Research in Education. London: Sage, 2012. 262 pp. ISBN: 9781446208076 (pbk).
- Babbie, E. (2007). Tile practice of social research. Istanbul Bilgi University Library.
- Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford university press.
- Bachman, L. F., & Palmer, A. S. (1979). Convergent and Discriminant Validation of Oral Language Proficiency Tests.
- Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests (Vol. 1). Oxford University Press.
- Bailey, K. M. (1996). The Best Laid Plans: Teachers' In-Class Decisions to Depart from Their Lesson Plans. In K. M. Bailey and D. Nunan (eds), Voices From the Language Classroom, pp. 15–40. Cambridge: Cambridge University Press.

- Balçikanlı, C. (2010). Learner autonomy in language learning: Student teachers' beliefs. *Australian Journal of Teacher Education (Online)*, 35(1), 90-103.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1-26.
- Bartels, N. (2009). Knowledge about language. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide to second language teacher education* (pp. 125-134). Cambridge University Press.
- Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, 25(2), 243–272.
- Başyurt Tüzel, A. E., & Akcan, S. (2009). Raising the language awareness of pre-service English teachers in an EFL context. *European Journal of Teacher Education*, 32(3), 271–287. doi:10.1080/02619760802572659.
- Baxter, P. & Jack, S. (1998). Qualitative case study methodology and implementation for novice teachers. *Qualitative Report*, 13, 544-559.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers?. *ELT Journal*, 73(2), 113-123.
- Bhattacharjee, A. (2012). *Social science research: Principles, methods, and practices*.
- Birello, M. (2012). Teacher cognition and language teacher education: beliefs and practice. A conversation with Simon Borg. *Bellaterra Journal of Teaching & Learning Language & Literature*, 5(2), 88-94.
- Bishop, C. (2001). *Case-based learning and the construction of professional practical knowledge in teacher education (Doctoral dissertation)*.
- Borg, M. (2001). Key concepts in ELT. Teachers' beliefs. *ELT Journal*, 55(2), 186–188. doi:10.1093/elt/55.2.186.
- Borg, S. (1999). Studying teacher cognition in second language grammar teaching. *System*, 27(1), 19-31.
- Borg, S. (2003a). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81–109. <https://doi.org/10.1017/s0261444803001903>.
- Borg, S. (2003b). Teacher cognition in grammar teaching. A literature review. *Language Awareness*, 12(2), 96-108..
- Borg, S. (2005). Experience, knowledge about language, and classroom experience in teaching grammar. In N. Bartels (Ed.), *Applied linguistics and language teacher education* (pp. 325-340). Springer.

- Borg, S. (2006). *Teacher Cognition and Language Education: Research and Practice*. Continuum.
- Borg, S. (2009). Language teacher cognition. In A. Burns & J. C. Richards (Eds.), *A Cambridge guide to second language teacher education* (pp. 163-171). Cambridge University Press.
- Borg, S. (2011). The impact of in-service teacher education on language teachers' beliefs. *system*, 39(3), 370-380.
- Borg, S. (2015). *Teacher Cognition and Language Education: Research and Practice*. Bloomsbury Publishing.
- Borg, S. (2019). Language teacher cognition: Perspectives and debates. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-319-58542-0_59-1
- Breen, M. P. (1991). Understanding the language teacher. In R. Phillipson, E. Kellerman, L. Selinker, M. Sharwood Smith & M. Swain (eds.), *Foreign/Second Language Pedagogy Research* (pp. 213–33). Clevedon, UK: Multilingual Matters.
- Breen, M. P., Hird, B., Milton, M., Oliver, R., & Thwaite, A. (2001). Making sense of language teaching: Teachers' principles and classroom practices. *Applied Linguistics*, 22(4), 470–501.
- Breitmayer, B. J. (1991). Triangulation in qualitative research: Issues of conceptual clarity and purpose. *Qual. Nurs. Res. Contemp. Dialogue*, 19, 226-239.
- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, 8(2), 153-169.
- Brookhart, S. M., & Freeman, D. J. (1992). Characteristics of entering teacher candidates. *Review of Educational Research*, 62(1), 37–60. <https://doi.org/10.2307/1170715>
- Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and Grading in Classrooms*. Pearson Prentice Hall.
- Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. Pearson Education.
- Bruner, J. (1996). *The Culture of Education*. Cambridge, MA: Harvard University Press.
- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.
- Bryman, A. (2014). June 1989 and beyond: Julia Brannen's contribution to mixed methods research. *International Journal of Social Research Methodology*, 17(2), 121-131.
- Burns, A. (1992). Teacher beliefs and their influence on classroom practice. *Prospect*, 7(3), 56-66.

- Burns, A. (1996). Starting all over again: From teaching adults to teaching beginners. In D. Freeman & J. C. Richards (eds.), *Teacher Learning in Language Teaching* (pp. 154–77). Cambridge: Cambridge University Press.
- 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.
- Cabaroglu, N., & Roberts, J. (2000). Development in student teachers' pre-existing beliefs during a 1-year PGCE programme. *System*, 28(5), 387-402. [https://doi.org/10.1016/S0346-251X\(00\)00019-1](https://doi.org/10.1016/S0346-251X(00)00019-1)
- Cabaroglu, N., & Yurdaisik, A. (2008). University instructors' views about and approaches to reading instruction and reading strategies. *Reading*, (8)2.
- Caffarella, R.S. & Merriam, S.B. (1999). Perspectives on adult learning: framing our research. In 40th Annual Adult Education Research Conference Proceedings. Northern Illinois University.
- Calderhead, J. (1996). Teachers: beliefs and knowledge. In D. C. Berliner & R. C. Calfee (eds.), *Handbook of Educational Psychology* (pp. 709–25). New York: Macmillan.
- Carr, E., & Ogle, D. (1987). KWL Plus: A Strategy for Comprehension and Summarization. *Journal of Reading*, 30, 626- 631.
- Carr, Michael. 2011. Critical analysis of statistical methods used in research papers. *MSOR Connections* 11 (2): 7-9.
- Çelik, S., & Coombe, C. (2021). Assessment literacy and assessment-literate EFL teachers: Why do we care? In S. Çelik, H. Çelik, & C. Coombe (Eds.), *Language assessment and test preparation in English as a foreign language (EFL) education* (pp. 1-13). Vizetek.
- Çelik, S., & Türkan, S. (2020). Assessment and evaluation in EFL: Classroom methods, standardized testing, and preparing learners for high-stakes exams. In S. Çelik (Ed.), *Approaches and principles in English as a foreign language (EFL) education* (3rd ed., pp. 421-428). Vizetek.
- Cengiz, C., & Karatas, F. O. (2015). Examining the effects of reflective journals on pre-service science teachers' general chemistry laboratory achievement. *Australian Journal of Teacher Education* (Online), 40(10), 125-146.
- Chan, Y. C. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English language teaching*, 7(1), 37-62.

- Chapelle, C. & Brindley, G. (2010). Assessment. In N. Schmitt (Ed.), *An introduction to applied linguistics: Second edition*. pp. 247-267. Hodder Education.
- Chappuis, J., & Stiggins, R. (2016). *An introduction to student-involved assessment for learning*. Pearson.
- Charmaz, K., & Thornberg, R. (2021). The pursuit of quality in grounded theory. *Qualitative research in psychology*, 18(3), 305-327.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389.
- Chilisa, B. & Kawulich, B. (2012). Selecting a research approach: Paradigm, methodology and methods. In C. Wagner, B. Kawulich, & M. Garner, (Eds.), *Doing Social Research: A Global Context* (pp. 51-61). McGraw-Hill Higher Education.
- Chisnall, P. M. (1996). Qualitative Interviewing: The Art of Hearing Data. *Journal of the Market Research Society*, 38(4), 553-555.
- Clark, C.M. and Peterson, P.L. (1984) *Teachers' thought Processes*. Institute for Research on Teaching, Michigan State University, East Lansing.
- Coffey, A., & Atkinson, P. (1996). *Making Sense of Qualitative Data, Complementary Research Strategies*. London, Thousand Oaks, CA and New Delhi: Sage Publications.
- Cohen, A. D. (1994). *Assessing language ability in the classroom* (2nd ed.). New York: Heinle and Heinle.
- Cohen, A. D., & Fass, L. (2001). Oral language instruction: Teacher and learner beliefs and the reality in EFL classes at a Colombian University. *Journal of Language and Culture*, 6, 43-62.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. (6th ed.) London: Routledge.
- Cohen, R. J., Swerdlik, M., & Sturman, E. (2012). *EBOOK: Psychological Testing and Assessment*. McGraw Hill.
- Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. London: Sage Publications.
- Creswell, J. W. (2007). *Qualitative Inquiry & Research Design Choosing Among Five Approaches* 2nd edition. London: Sage Publications.
- Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd Ed.). Thousand Oaks, CA: Sage.

- Creswell, J. W. (2014). The selection of a research approach. *Research design: Qualitative, quantitative, and mixed methods approaches*, (pp.3-24). Thousand Oaks: Sage Publications.
- Creswell, J.W. (2009). *Qualitative, quantitative and mixed methods approaches*, London: Sage Publications.
- Creswell, J.W. and Poth, C.N. (2018) *Qualitative Inquiry and Research Design Choosing among Five Approaches*. 4th Edition, SAGE Publications, Inc., Thousand Oaks.
- Cross, R. (2010). Language teaching as sociocultural activity: Rethinking language
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7, 31-51.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334.
- 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. <https://doi.org/10.1080/0969594X.2010.516643>
- DeLuca, C., Luu, K., Sun, Y., & Klinger, D. A. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters*, 4, 5-29.
- DeLuca, C., Valiquette, A., & Klinger, D. A. (2016). Implementing Assessment for Learning in Canada: The challenge of teacher professional development. In *Assessment for learning: Meeting the challenge of implementation* (pp. 145-160). Springer, Cham.
- Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA: D.C. Heath & Co Publishers.
- Diener, E., & Crandall, R. (1978). *Ethics in social and behavioral research*. U Chicago Press.
- DiRanna, K., Osmundson, E., Topps, J., Barakos, L., Cerwin, K., & Carnahan, D. (2008). *Assessment-centered teaching: A reflective practice*. Corwin Press.
- Dobson, R.L., Dobson, J.E., 1983. Teacher beliefs-practice congruency. *Viewpoints in Teaching and Learning* 59 (1), 20–27.
- Doolittle, P. E. (2003). *Constructivism and Online Education* (Doolittle). Retrieved March, 5th, 2018 from <https://jgregorymcverry.com/>

- Dornyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.
- Dornyei, Z. and Ushioda, E. (2011) *Teaching and Researching Motivation*. 2nd Edition, Pearson, Harlow.
- Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching: Motivation*. Routledge.
- Driscoll, M. (2000). *Psychology of Learning for Instruction*. Boston: Allyn& Bacon.
- Duberley, J., Johnson, P., & Cassell, C. (2012). Philosophies underpinning qualitative research. *Qualitative organizational research: Core methods and current challenges*, 15.
- Duckworth, E. (1987). *The having of wonderful ideas and essays on teaching and learning*. New York: Teacher College Press
- Elbaz, F. (1981). The teacher's 'practical knowledge': Report of a case study. *Curriculum Inquiry*, 11(1), 43–71.
- England, N. (2017). Developing an interpretation of collective beliefs in language teacher cognition research. *TESOL Quarterly*, 51,229–238. <https://doi.org/10.1002/tesq.334>.
- Erickson, F. (1986). *Qualitative Methods in Research on Teaching*. In M. Wittrockk (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 119-161). New York: MacMillan.
- Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher Technology Change. *Journal of Research on Technology in Education*, 42(3), 255–284. doi:10.1080/15391523.2010.10782551
- Ezzy, D. (2002) *Qualitative Analysis. Practice and innovation*. London: Routledge.
- Falsgraf, C. (2005, April). Why a national assessment summit? New visions in action. National Assessment Summit. Meeting conducted in Alexandria, Va. Retrieved from: http://www.nflrc.iastate.edu/nva/word_documents/assessment_2005/pdf/nsap_introduction.pdf.
- Farhady, H. (2012). Principles of language assesment. In C. Coombe, P. Davidson, B. O'Sullivan, and S. Stoyhoff. (Eds.). *The Cambridge guide to second language assessment*. (pp. 37-46). Cambridge University Press, USA.
- Farrell, T. S. C. (2008). *Teaching Reading to English Language Learners: A reflective guide*. Thousand Oaks, CA: Corwin Press.
- Farrell, T.S.C., and L. P. C. Patricia. (2005). Conceptions of Grammar Teaching: A case study of Teachers' Beliefs and Classroom Practices. *Teaching English as a second language-e-Journal*, 9(2).

- Feryok, A. (2012). Activity Theory and language teacher agency. *The Modern Language Journal*, 96(1) 95-107. <https://doi.org/10.1111/j.1540-4781.2012.01279.x>
- Fitzpatrick, J. J. and Wallace, M. (2006). *Encyclopedia of nursing research*. New York: Springer Publishing.
- Flick, U. (2009). *An Introduction to Qualitative Research* (4th ed.). London: SAGE Publications Ltd.
- Flores, B. B. (2001). Bilingual education teachers' beliefs and their relation to self-reported practices. *Bilingual Research Journal*, 25, 251-275.
- Foss, D. H., & Kleinsasser, R. C., (1996). Preservice elementary teachers' views of pedagogical and mathematical content knowledge. *Teaching and Teacher Education*, 12, 429-442.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to Design and Evaluate Research in Education*. New York: Mc. Graw Hill
- Frankel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. New York: Pearson.
- Freeman, D. (1992). Language teacher education, emerging discourse, and change in classroom practice. In J. Flowerdew, M. Brock, & S. Hsia (Eds.), *Perspectives in second language teacher education* (pp. 1-21). Hong Kong: City Polytechnic of Hong Kong.
- Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach. A perspective from north American educational research on teacher education in English language teaching. *Language Teaching*, 35, 1-13. doi:10.1017/S0261444801001720
- Freeman, D., & Johnson, K. E. (2005). Toward linking teacher knowledge and student learning. In D.J. Tedick (Ed.), *Second language teacher education: International Perspective*. (pp.73-95). Mahwah: NJ Lawrence Erlbaum Associates.”
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>.
- Gabillon, Z. (2012). Discrepancies between L2 Teacher and L2 Learner Beliefs. *English Language Teaching*, 5(12). doi:10.5539/elt.v5n12p94.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston: Allyn & Bacon.
- Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language learning*, 55(4), 575-611.

- Gatbonton, E. (1999). Investigating experienced ESL teachers' pedagogical knowledge. *The Modern Language Journal*, 83 (1), 35–50.
- Gatbonton, E. (2008). Looking beyond teachers' classroom behaviour: Novice and experienced ESL teachers' pedagogical knowledge. *Language teaching research*, 12(2), 161-182.
- Gholami, K. and J. Husu (2010). How do teachers reason about their practice? Representing the epistemic nature of teachers' practical knowledge. *Teaching and Teacher Education*, 26, 1520-1529.
- Gillham, B. (2001). *Research interview*. Bloomsbury Publishing.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>
- Goker, S.D. 2006. Impact of peer coaching on self-efficacy and instructional skills in TEFL teacher education. *System*34: 239–54.
- Golombek, P. R. (1998). A study of language teachers' personal practical knowledge. *TESOL Quarterly*, 32(3), 447–464.
- 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.
- Gray, A. J. (1997). *Constructivist teaching and learning* (pp. 97-07). Saskatchewan School Trustees Association.
- Green, J., Franquiz, M., & Dixon, C. (1997). The myth of the objective transcript: Transcribing as a situated act. *Tesol Quarterly*, 31(1), 172-176.
- Grisham, D. L. (2000). Connecting theoretical conceptions of reading to practice: a longitudinal study of elementary school teachers. *Reading Psychology*, 21(2), 145–170. doi:10.1080/02702710050084464
- Gronlund, N. E. (1998). *Assessment of student achievement* (6th edition). Boston: Allyn and Bacon
- Grossman, P., & Richert, R. (1988). Unacknowledged knowledge growth: A reexamination of the effects of teacher education. *Teaching and Teacher Education*, 4, 53-62. doi:10.1016/0742-051X(88)90024-8
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163-194), 105.

- Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education & Literacy Studies*, 3(4), 42-48.
- Hall, L. A. (2005). Teachers and content area reading: Attitudes, beliefs and change. *Teaching and Teacher Education*, 21(4), 403–414. doi:10.1016/j.tate.2005.01.009
- Harlen, W. (2005). Teachers' summative practices and assessment for learning—tensions and synergies. *Curriculum Journal*, 16(2), 207-223.
- Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. Suny Press.
- 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.
- Haight, J., & Crusan, D. (2016). Filling the gaps: L2 grammar and assessment preparation for ELA teachers. In *Teaching English language arts to English language learners* (pp. 171-192). Palgrave Macmillan, London.
- Henning, G. (1987). *A guide to language testing, development, evaluation and research*. Cambridge, MA: Newbury House.
- Herrera, L., & Macias, D. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. doi:10.14483/udistrital.jour.calj.2015.2.a09.
- Hislam, J., & Cajkler, W. (2005). Teacher trainees' explicit knowledge of grammar and primary curriculum requirements in England. In *Applied linguistics and language teacher education* (pp. 295-312). Springer, Boston, MA.
- Hoffman, J. V., & Kugle, C. L. (1982). A study of theoretical orientation to reading and its relationship to teacher verbal feedback during reading instruction. *Journal of Classroom Interaction*, 18(1), 2–7.
- Holmes G. 2007. Protection, politics and protest: Understanding resistance to conservation. *Conservation and Society* 5:184–201
- Holt-Reynolds, D. (1992). Personal history-based beliefs as relevant prior knowledge in course work. *American Educational Research Journal*, 29(2), 325-349. <https://doi.org/10.3102/00028312029002325>.
- Hook, C. M., & Rosenshine, B. V. (1979). Accuracy of teacher reports of their classroom behavior. *Review of Educational Research*, 49(1), 1–11. <https://doi.org/10.2307/1169924>.

- Howerton, A. M. (2016). Elephant on a stepladder: an exploration of pre-service English teacher assessment literacy [Doctoral dissertation, Northern Illinois University]. Northern Illinois University Institutional Repository. <https://commons.lib.niu.edu/handle/10843/20902>
- Huang, J., & He, Z. (2016). Exploring assessment literacy. *Higher Education of Social Science*, 11(2), 18-27. <http://dx.doi.org/10.3968/8727>
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Ilin, G. (2003). A model for effective supervision from the supervisor and the studentteachers' perspective: A social constructivist approach. Unpublished Ph.D Thesis, Çukurova University, Adana.
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapella (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-9). Blackwell Publishing. <https://doi.org/10.1002/9781405198431.wbeal0605>
- Inbar-Lourie, O. (2013b). Guest editorial to the special issue on language assessment literacy. *Language Testing*, 30(3), 301–307.
- Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, and S. May (Eds), *Language Testing and Assessment* (pp. 1-14). Springer International Publishing.
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *Educational Research Association The International Journal of Research in Teacher Education*. 6(2): 26-37.
- Jin Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555–584.
- Johnson, K. E. (1992). The Relationship between Teachers' Beliefs and Practices during Literacy Instruction for Non-Native Speakers of English. *Journal of Reading Behavior*, 24(1), 83-108. <https://doi.org/10.1080/10862969209547763>.
- Johnson, K. E. (1994). The Emerging Beliefs and Instructional Practices of Pre-Service English as Second Language Teachers. *Teaching and Teacher Education*, 10(4), 439-452. [https://doi.org/10.1016/0742-051X\(94\)90024-8](https://doi.org/10.1016/0742-051X(94)90024-8)
- Johnson, K. E. (1996). The vision versus the reality: The tensions of the TESOL practicum. In D. Freeman & J. C. Richards (eds.), *Teacher Learning in Language Teaching* (pp. 30–49). Cambridge: Cambridge University Press.
- Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. New York: The Routledge.

- Johnson, K. E. (2018). Studying language teacher cognition: Understanding and enacting theoretically consistent instructional practices. *Language Teaching Research*, 22(3) 259-263.
- Johnston, B., & Goettsch, K. (2000). In search of the knowledge base of language teaching: Explanations by experienced teachers. *Canadian Modern Language Review*, 56, 437-468.
- Kagan, D. M. (1992a). Implication of Research on Teacher Belief. *Educational Psychologist*, 27(1), 65–90. doi:10.1207/s15326985ep2701_6.
- Kagan, D. M. (1992b). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62, 129–69.
- Kaminski, T. W., & Pitney, W. A. (2004). Strategies for Establishing Trustworthiness in Qualitative Research. *Athletic Therapy Today*, 9(1), 26–28. doi:10.1123/att.9.1.26.
- Katz, A. (2012). Linking assessment with instructional aims and learning. *The Cambridge guide to second language assessment*, 66-73.
- Khadijeh, B., & Amir, R. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, 3(1), 139-146. <https://doi.org/10.5296/ijele.v3i1.6887>.
- Kiss, T., & Lin, C. (2016). Beliefs, knowledge and practices of grammar pedagogy: Teaching dyslexic learners. *Malaysian Journal of ELT Research*, 12(1), 37-59. https://melta.org.my/journals/MAJER/downloads/majer12_01_03.pdf.
- Krathwohl, D. R. (1998). *Methods of educational and social science research* (2nd ed.). New York: Addison Wesley Longman.
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, 30(3), 411-433.
- Kubanyiova, M. (2015). The role of teachers' future self-guides in creating L2 development opportunities in teacher-led classroom discourse: Reclaiming the relevance of language teacher cognition. *The Modern Language Journal*, 99(3), 565-584. <https://doi.org/10.1111/modl.12244>.
- Kubanyiova, M., & Feryok, A. (2015). Language teacher cognition in applied linguistics research: Revisiting the territory, redrawing the boundaries, reclaiming the relevance. *The Modern Language Journal*, 99(3), 435-449.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32 (2), 169-197.

- Lam, Y. (2000). Technophilia vs. technophobia: A preliminary look at why second-language teachers do or do not use technology in their classrooms. *Canadian Modern Language Review*, 56 (3), 390–420.
- Lantolf, J. & Thorne, S.L. (2007). Sociocultural theory and second language learning. In B. van Patten & J. Williams (Eds.). *Theories*
- Lantolf, J. P. (2001). Sociocultural theory and SLA. In R. B. Kaplan (Ed.), *Handbook of applied linguistics* (pp. 109– 119). Oxford: Oxford University Press.
- Lantolf, J. P. (2006). Sociocultural theory and L2: State of the art. *Studies in Second Language Acquisition*, 28 (1), 67-109.
- Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. Oxford: Oxford University Press.
- Li, L. (2008). EFL teachers' beliefs about ICT integration in Chinese secondary schools [Unpublished doctoral thesis]. Queen's University, Belfast.
- Li, L. (2013). The complexity of language teachers' beliefs and practice: one EFL teacher's theories. *Language Learning Journal*, 41(2), 175–191.
- Li, L. (2017). *Social interaction and teacher cognition*. Edinburgh University Press.
- Li, L. (2020). *Language teacher cognition: A sociocultural perspective*. Palgrave Macmillan.
- Li, S. (2012). The effects of input-based practice on pragmatic development of requests in L2 Chinese. *Language Learning*, 62(2), 403-438.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. sage.
- Lortie, D. (1975). *Schoolteacher: A Sociological Study*. London: University of Chicago Press.
- Loughran, J., & Russell, T. (Eds.) (1997). *Purpose, passion and pedagogy in teacher education*. London: Falmer Press.
- Lune, H., & Berg, B. L. (2017). *Qualitative research methods for the social sciences*. Pearson.
- Maloch, B., Flint, A., Eldridge, D., Harmon, J., Loven, R., Fine, J. , Martinez, M. (2003). Understandings, beliefs, and reported decision making of first-year teachers from different reading teacher preparation programs. *Elementary School Journal*, 103, 431–457.
- Malone, M. E. (2008). Training in language assessment. In E. Shohamy & N. H. Hornberger (Eds.), *The Encyclopedia of Language and Education* (2nd ed., Vol. 7) (pp. 225-239). Blackwell Publishing.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. doi:10.1177/0265532213480129

- Mangubhai, F., Marland, P., Dashwood, A., & Son, J. B. (2004). Teaching a foreign language: One teacher's practical theory. *Teaching and Teacher Education*, 20(3), 291-311.
- Marshall, C. and Rossman, G. B. (1999) *Designing Qualitative Research*, 3rd edn. Thousand Oaks, CA: SAGE.
- Mattheoudakis, M. (2007). Tracking changes in pre-service EFL teacher beliefs in Greece: A longitudinal study. *Teaching and teacher education*, 23(8), 1272-1288.
- Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. London: Falmer Press.
- McDowell, L., Sambell, K., Bazin, V., Penlington, R., Wakelin, D., Wickes, H., & Smailes, J. (2006). *Assessment for learning: Current practice exemplars from the centre for excellence in teaching and learning in assessment for learning*. Northumbria University.
- McMillan, J. H. & Schumacher, S. (1993). *Research in education: A conceptual introduction* (3rd ed.). New York: Harper Collins.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context.
- Meijer, P. C., Verloop, N., & Beijaard, D. (1999). Exploring language teachers' practical knowledge about teaching reading comprehension. *Teaching and Teacher Education*, 15, 59-84.
- Menard, S. (2002). *Longitudinal research* (Vol. 76). Sage.
- Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education*. Revised and Expanded from "Case Study Research in Education.". Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104.
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Mertler, C. A. (2003). Preservice Versus Inservice Teachers' Assessment Literacy: Does Classroom Experience Make a Difference?.
- Mertler, C. A. (2009a). Teachers' assessment knowledge and the perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12 (2), 101-113 DOI: 10.1177/1365480209105575.
- Mertler, C. A. (2009b). *Action research: Teachers as researchers in the classroom*. Sage.
- Mertler, C. A., & Campbell, C. (2005). *Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory"*. Online Submission.

- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Miles, M.B., Huberman, A.M. and Saldana, J. (2014) *Qualitative Data Analysis: A Methods Sourcebook*. Sage, London.
- Moodie, I. (2016). The anti-apprenticeship of observation: how negative prior language learning experience influences English language teachers' beliefs and practices. *System*, 60, 29–41. <https://doi.org/10.1016/j.system.2016.05.011>
- Morina, M. (2016). *Teachers' Beliefs and Practices about Grammar Teaching in Public High Schools in Kosovo*. Graduate Theses and Dissertations Retrieved from <https://scholarworks.uark.edu/etd/1620>
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile: Issues in Teachers' Professional Development*, 14(1).
- Ngo, X.M. (2018). A sociocultural perspective on second language writing teacher cognition: A Vietnamese teacher's narrative. *System*, 78, 79-90.
- Nishino, T. (2008). Japanese secondary school teachers' beliefs and practices regarding communicative language teaching: An exploratory survey. *JALT journal*, 30(1), 27.
- Norris, J. M. (2006). The issue: The why (and how) of assessing student learning outcomes in college foreign language programs. *The Modern Language Journal*, 90(4), 576-583.
- Numrich, C. (1996). On Becoming a Language Teacher: Insights from Diary Studies. *TESOL Quarterly*, 30(1), 131-153. <https://doi.org/10.2307/3587610>
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30 (3), 363-380.
- Ogle, D. M. (1986). KWL: A teaching model that develops active reading of expository text. *The reading teacher*, 39(6), 564-570.
- Ölmezer-Öztürk, E. (2018). *Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers*. Unpublished PhD Dissertation, Anadolu University, Turkey.
- Ölmezer-Öztürk, E., & Aydin, B. (2018). Toward measuring language teachers' assessment knowledge: development and validation of Language Assessment Knowledge Scale (LAKS). *Language Testing in Asia*, 8(1), 1-15.
- Önalın, O., & Karagül, 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.

- Ostovar-Nameghi, S. A., & Sheikahmadi, M. (2016). From Teacher Isolation to Teacher Collaboration: Theoretical Perspectives and Empirical Findings. *English Language Teaching*, 9(5), 197-205.
- Öz, S. & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Özmen, K. S. (2012). Exploring student teachers' beliefs about language learning and teaching: A longitudinal study. *Current Issues in Education*, 15(1).
- Öztürk, G. (2015). Language teacher cognition, classroom practices and institutional context: A qualitative case study on three EFL teachers [Unpublished doctoral dissertation]. Middle East Technical University.
- Öztürk, M. (2014). EFL instructors' cognitions and actions in relation to foreign language learning and teaching processes [Unpublished doctoral dissertation]. Middle East Technical University.
- Pajares, M.F. (1992) Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, 62, 307-332. <http://dx.doi.org/10.3102/00346543062003307>
- Palomba, C. A., & Banta, T. W. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. Higher and Adult Education Series. Jossey-Bass, Inc., Publishers, 350 Sansome Street, San Francisco, CA 94104.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation* (No. 4). Sage.
- Patton, M. Q. (2005). Qualitative research. *Encyclopedia of statistics in behavioral science*.
- Peacock, M. (1999). Beliefs about language learning and their relationship to proficiency. *International Journal of Applied Linguistics*, 9(2), 247–263.
- Peacock, M. (2001). Pre-service ESL teachers' beliefs about second language learning: a longitudinal study. *System*, 29, 177–95.
- Pearson, J. Are teachers' beliefs incongruent with their observed classroom behavior?. *Urban Rev* 17, 128–146 (1985). <https://doi.org/10.1007/BF01108253>
- Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37(3), 380–390. doi:10.1016/j.system.2009.03.002.
- Piaget, J. (1972). *The psychology of the child*. New York: Basic Books.
- Pitsoe, V. J., & Maila, W. M. (2012). Towards constructivist teacher professional development. *Journal of Social Sciences*, 8(3), 318-324.
- Plake, B. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Midwestern Educational Researcher* 6(1), 21-27 Popham, W.

- J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, 48(1), 4–11. doi:10.1080/00405840802577536
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12.
- Popham, W. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273.
- Popham, W. J. (2006). *Mastering assessment: A selfservice system for educators*. New York: Routledge.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48, 4–11.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249-258.
- Richards, J. and Lockhart, C. (1994) *Reflective Teaching in Second Language Classrooms*. Foreign Language Teaching and Research Press, Cambridge University Press. <https://doi.org/10.1017/CBO9780511667169>.
- Richards, J. C., Gallo, P. B., & Renandya, W. A. (2001). Exploring teachers' beliefs and the processes of change. *PAC journal*, 1(1), 41-58.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (ed.), *Handbook of Research on Teacher Education* (2nd edn., pp. 102–19). New York: Macmillan.
- Riessman, C. K. (1993). *Narrative analysis* (Vol. 30). Sage.
- Rock, T. C. & Wilson, C. (2005). Improving teaching through lesson study. *Teacher Education Quarterly*, 32(1), 77-92.
- Rodgers, C. R. (2002). Seeing student learning: Teacher change and the role of reflection. *Harvard educational review*, 72(2), 230.
- Rogier, D. (2014). *Assessment Literacy: Building a Base for Better Teaching and Learning*. In *English Teaching Forum* (Vol. 52, No. 3, pp. 2-13). US Department of State. Bureau of Educational and Cultural Affairs, Office of English Language Programs, SA-5, 2200 C Street NW 4th Floor, Washington, DC 20037.
- Saldaña, J. (2009). *The coding manual for qualitative researchers*. Sage Publications Ltd.
- Saldaña, J., & Omasta, M. (2016). *Qualitative research: Analyzing life*. Sage Publications.
- Sato, K., & Kleinsasser, R. C. (1999). Communicative language teaching (CLT): Practical understandings. *The Modern Language Journal*, 83(4), 494-517.

- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327.
- Schön, D. A. (1987). *Educating the reflective practitioner*. Washington, DC: American Educational Research Association.
- Schon, D.A. (1983) *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, New York.
- Schwandt, T. A. (2001). Dictionary of qualitative inquiry. In *Dictionary of qualitative inquiry* (pp. xxxiv-281).
- Seferoğlu, G., S. Korkmazgil, and Z. Ölçü. (2009). Gaining insights into teachers' ways of thinking via metaphors. *Educational Studies*, 35(3), 323-335.
- Sendan, F. & J. Roberts (1998). Orhan: a case study in the development of a student teachers' personal theories. *Teachers and Teaching: Theory and Practice*, 4, 229-44.
- Sert, O. (2015). *Social interaction and L2 classroom discourse*. Edinburgh University Press.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, 22(2), 63-75.
- Sherman, R. R., & Webb, R. B. (Eds.). (1988). *Qualitative research in education: Focus and methods* (Vol. 3). Psychology Press
- Shermila, A. J. (2013). *Constructivist Teaching and Learning*.
- Shim, K. N. (2009). *An Investigation into teachers' perceptions of classroom-based assessment of English as a foreign language in Korean primary education*. [Doctoral dissertation], University of Exeter.
- Smith, D. B. (1996). Teacher decision making in the adult ESL classroom. In D. Freeman & J. C. Richards (eds.), *Teacher Learning in Language Teaching* (pp. 197-216). Cambridge: Cambridge University Press.
- Solak, E., & Semerci, H. (2015). İkinci Sınıfta İngilizce Öğretimine Yönelik Öğretmen Ve Öğrenci Görüşleri. *Dil ve Edebiyat Eğitimi Dergisi*, (16).
- Stake, R. E. (1995). *The art of case study research*. sage.
- Stiggins, R. (2006). *Balanced assessment systems: Redefining excellence in assessment*. Educational Testing Service, 1-10.
- Stiggins, R. (2007). Assessment through the student's eyes. *Educational leadership*, 64(8), 22.
- Stiggins, R. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 233-250). Routledge, Taylor and Francis Group.

- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-39.
- Stiggins, R. J. (1999). Assessment, student confidence, and school success. *The Phi Delta Kappan*, 81(3), 191-198.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.
- Stiggins, R. J. (2010). "Essential formative assessment competencies for teachers and school leaders," in *Handbook of Formative Assessment*, eds H. L. Andrade and G. J. Cizek (New York, NY: Routledge, 233–250.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1-14.
- Tabachnick, B.R., Zeichner, K.M., 1986. Teacher beliefs and classroom behaviours: some teacher responses to inconsistency. In: BenPeretz, M., Bromme, R., Halkes, R. (Eds.), *Advances of Research on Teacher Thinking*. Swets and Zeitlinger, Lisse, Netherlands, pp. 84–96.
- Tao, N. (2014). Development and validation of classroom assessment literacy scales: English as a Foreign Language (EFL) teachers in a Cambodian Higher Education Setting. Unpublished PhD dissertation, Victoria University, Australia.
- Taras, M. (2005). Assessment—summative and formative—some theoretical reflections. *British journal of educational studies*, 53(4), 466-478.
- Tavakoli, H. (2012). *A dictionary of research methodology and statistics in applied linguistics*. Rahnama press.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36.
- teacher practice. *Modern Language Journal*, 94(3), 434-452.
- Tercanlioğlu, L., 2001. Pre-service teachers as readers and future teachers of EFL reading. *TESL-EJ* [online], 5(3). Available from: <http://www-writing.berkeley.edu/TESL-EJ/ej19/a2.html> [Accessed 14 September 2007].
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Tsang, W. K. (2004). Teachers' personal practical knowledge and interactive decisions. *Language teaching research*, 8(2), 163-198.

- Tsui, A. B. M. (1996). Learning how to teach ESL writing. In D. Freeman & J. C. Richards (Eds.), *Teacher Learning in Language Teaching* (pp. 97-119). Cambridge University Press.
- Tsui, A. B. M. (2011). Teacher Education and Teacher Development. In E. Hinkel (ed). *Handbook of Research in Second Language Teaching and Learning: Volume II*. Routledge. (pp. 21-39).
- Tsui, A.B.M. (2003). *Understanding expertise in teaching: Case studies of ESL teachers*. Cambridge: Cambridge University Press.
- Urmston, A. (2003). Learning to teach English in Hong Kong: The opinions of teachers in training. *Language and Education*. 17(2). 112-137.
- Üstünel, E. (2008). The relationship between trainee teachers' views and practices related to classroom discipline. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*. 35, 332-341.
- Van der Vleuten, C. P. M., Norman, G. R., & De Graaff, E. (1991). Pitfalls in the pursuit of objectivity: issues of reliability. *Medical education*, 25(2), 110-118.
- Van Lier, L. (2005). Case Study. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 195-208). Mahwah, NJ: Erlbaum.
- Vogt, K., & Tzagari, 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 Kusen, T. and Poslethwaite, N. (Eds) *International Encyclopedia of Education* (supplementary volume 162–163). Oxford: Pergamon.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wach, A. (2012). Classroom-based language efficiency assessment: A challenge for EFL teachers. *Glottodidactica*, 39(1), 81-92.
- Walberg, H. (1972). Decision and perception: New constructs in research on teaching effects. *Cambridge Journal of Education*, 7, 12–20.
- Wall, D. (2012). Washback. In G. Fulcher and D. Fulcher (Eds.). *Routledge handbook of language testing* (pp. 79-92). London and New York: Routledge.

- Wang, P. (2011). Constructivism and learner autonomy in foreign language teaching and learning: To what extent does theory inform practice. *Theory and Practice in Language Studies*, 1(3), 273-277.
- Warford, M. K. (2011). The zone of proximal teacher development. *Teaching and teacher education*, 27(2), 252-258.
- Warford, M. K., & Reeves, J. (2003). 'Falling into it': Novice TESOL teacher thinking. *Teachers and Teaching: Theory and Practice*, 9, 47e65.
- Webb, N. L. (2002). Assessment literacy in a standards-based urban education setting. A paper presented at the American Educational Research Association Annual Meeting. New Orleans, Louisiana. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.573.676&rep=rep1&type=pdf>
- Weber, R. P. (1990). *Basic content analysis* (Vol. 49). Sage.
- White, B. (2009). Student perceptions of what makes good teaching. Refereed paper presented at 'Teacher education crossing borders: Cultures, contexts, communities and curriculum' the annual conference of the Australian Teacher Education Association (ATEA), Albury, 28 June – 1 July.
- Wickham, M., & Woods, M. (2005). Reflecting on the strategic use of CAQDAS to manage and report on the qualitative research process. *The qualitative report*, 10(4), 687-702.
- William, M., & Burden, R. L. (1997). *Psychology for language teachers: a social constructivist approach*. Cambridge: Cambridge University Press.
- Winegar, L.T. (1997). Can internalization be more than a magical phrase? Notes toward the constructive negotiation of this process. In C. Lightfoot & B.D. Cox (Eds.). *Sociogenetic perspectives on internalization* (pp.25-43). Mahwah, NJ: Erlbaum.
- Wood, R. 1993. *Assessment and Testing*. Cambridge: Cambridge University Press.
- Woods, D., & Çakır, H. (2011). Two dimensions of teacher knowledge: The case of communicative language teaching. *System*, 39(3), 381-390. doi:10.1016/j.system.2011.07.010
- Wray, D. (1993). Student-Teachers' Knowledge and Beliefs about Language. In N. Bennett, & C. Carré (Eds.), *Learning to Teach* (pp. 51-72). London: Routledge.
- Xie, Q., and Tan, S. (2019). Preparing primary English teachers in Hong Kong: focusing on language assessment literacy. *J. Asiat.* 16, 653–673. doi: 10.18823/asiatefl.2019.16.2.14.653
- Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162.

- Y. Zhang and B. M. Wildemuth, "Qualitative Analysis of Content," In: B. M. Wildemuth, Ed., *Applications of Social Research Methods to Questions in Information and Library Science*, Libraries Unlimited, 2009. pp. 1-12.
- Yıldırım, A. & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri*. (9. Baskı). Ankara:SeçkinYayıncılık.
- Yin, M. (2010). Understanding classroom language assessment through teacher thinking research. *Lang. Assess. Q.* 7, 175–194. doi: 10.1080/15434300903447736
- Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Yin, R. K. (2014). *Case study research: Design and methods*. Los Angeles, CA: Sage.
- Yook, C. M. (2010). *Korean Teachers' Beliefs about English Language Education and their Impacts upon the Ministry of Education-Initiated Reforms*. Dissertation, Georgia State University, 2010..
- Zaidah, Z. (2007) *Case Study as a Research Method*. *Jurnal Kemanusiaan*, 9, 1-6.
- Zeichner, K. M., & Liston, D. P. (1996). *Reflective teaching: An introduction*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied measurement in education*, 16(4), 323-342.

APPENDIX B. KWL CHART

Topic:

Date: ____ / ____ / ____

K – W – L CHART

What I know	W I wonder (or want to learn)	What I have learnt

PS: You can expand the boxes as much as you need.

APPENDIX C. SPECIFIC DATES FOR OVERALL RESEARCH PROCEDURE

2-8 September, 2019	1st individual pre semi-structured interviews
9-15 September, 2019	1st classroom observations and follow-up semi-structured interviews
16-22 September, 2019	2nd individual pre semi-structured interviews
23-30 September, 2019	Classroom observations and follow-up interviews
30 September- 6 October, 2019	1st RRD session
7-19 October, 2019	2nd RRD session
19-29 October, 2019	3 RRD session
29 October-9th November, 2019	4 RRD session
9-19 November, 2019:	5 RRD session
19-30 November, 2019	6 RRD session
1-11 December, 2019	7 RRD session
11-21 December, 2019	8 RRD session
22 December, 2019 – 1 January, 2020	9 RRD session
2-12 January, 2020	10 RRD session
13 – 20 January, 2020	Collection of Final KWL charts and reflection reports
20 – 26 January, 2020	1st last individual semi-structured interviews
27 January – 2 February, 2020	2nd last individual semi-structured interviews
03 – 15 February, 2020	1st and second classroom observations
16 - 29 February, 2020	Follow up interviews

APPENDIX D. READING LIST FOR EACH REFLECTIVE READING AND DISCUSSION SESSION

Week	Topic	Target reading list for the week
1	Definitions of terms such as testing, assessment, measurement, evaluation. Formative vs Summative assessment	<p>Investigating Formative Assessment: Teaching, Learning, and Assessment in the Classroom (Torrance & Tatlor)</p> <p>Fundamental considerations in language testing, Lyle F. Bachman Dante D. Dixson & Frank C. Worrell (2016) Formative and Summative Assessment in the Classroom, <i>Theory Into Practice</i>, 55:2, 153-159.</p> <p>Qu, Wenjie, and Chunling Zhang. "The analysis of summative assessment and formative assessment and their roles in College English assessment system." <i>Journal of Language Teaching and Research</i>, vol.4, no.2, 2013</p> <p>William, D. (2011). What is assessment for learning. <i>Studies in Educational Evaluation</i> 37(2011) 3-14.</p> <p>Handbook of Assessment for Language Teachers (Dina Tsagari, et al)</p> <p>Measurement and Assessment in Teaching (Miller et al)</p> <p>Language Assessment Principles and Classroom Practices (Douglas Brown)</p> <p>Handbook for the English Classroom: Assessment (Harris, M. & McCann, p)</p> <p>Classroom Assessment in Action (Shermis & Vasta)</p> <p>Classroom Assessment Concepts and Applications (Russel & Airasion)</p>
2	Beyond the tests: Alternatives in language assessment	<p>Developing Effective Assessment in Higher Education: a practical guide (Bloxhom & Boyd)</p> <p>Classroom Assessment in Action (Shermis & Vasta)</p> <p>Assessment in the Classroom: Constructing and Interpreting the tests (George, K. Cunningham)</p> <p>Handbook of Assessment for Language Teachers (Dina Tsagari, et al)</p> <p>Handbook for the English Classroom: Assessment (Harris, M. & McCann, p)</p> <p>Practical Language Testing (Glenn Fulcher)</p> <p>Measurement and Assessment in Teaching (Miller et al)</p> <p>Assessment and Learning in Secondary School, 2004.</p>
3	Developing valid grading procedures for assessing and evaluating students	<p>Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson)</p> <p>Language Assessment Principles and Classroom Practices (Douglas Brown)</p> <p>Assessment in the Classroom: Constructing and Interpreting the tests (George, K. Cunningham)</p> <p>Classroom Assessment in Action (Shermis & Vasta)</p> <p>Classroom Assessment Concepts and Applications (Russel & Airasion)</p>

		Classroom Assessment: Enhancing the Quality of Teacher Decision Making (Anderson, L., W)
4	Using assessment results to make decisions about students, planning education, developing curriculum, and school improvement.	Assessment in the Classroom: Constructing and Interpreting the tests (George, K. Cunningham) Classroom Assessment Concepts and Applications (Russel & Airasion) Mehrens, W., A. & Lehmann, I., J.) Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson)
5	Communicating assessment results to students and other stakeholders.	Classroom Assessment Concepts and Applications (Russel & Airasion) Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson)
6	Recognizing any unethical and illegal assessment (methods) and evaluation practices	Classroom Assessment in Action (Shermis & Vasta) Estajhi, M. (2011). Ethics and Validity Stance in Educational Assessment. <i>English Language and Literature Studies</i> 1(2), 89-99. Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson) Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: issues and attitudes. <i>Teaching and Teacher Education</i> , 23(7), 999–1011. Assessment in the Classroom: Constructing and Interpreting the tests (George, K. Cunningham) Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson)
7	Administering exams, scoring them and interpreting the results of both externally produced and teacher produced assessment methods.	Mehrens, W., A. & Lehmann, I., J.) Classroom Assessment Concepts and Applications (Russel & Airasion) Practical Language Testing (Glenn Fulcher) Measurement and Assessment in Teaching (Miller et al) Language Assessment Principles and Classroom Practices (Douglas Brown) Assessment in the Classroom: Constructing and Interpreting the tests (George, K. Cunningham) Classroom Assessment Concepts and Applications (Russel & Airasion) Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson)
8	Objective vs Subjective tests. Writing the test items	Measurement and Evaluation in Education and Psychology (Mehrens, W., A. & Lehmann, I., J.) Measurement and Assessment in Teaching (Miller et al) Assessment in the Classroom: Constructing and Interpreting the tests (George, K. Cunningham) Classroom Assessment in Action (Shermis & Vasta) Classroom Assessment Concepts and Applications (Russel & Airasion) Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson)

9	Choosing / Developing assessment methods according to instructional purposes. Testing the tests: Principles for evaluating / designing tests (Practicability, authenticity, validity, reliability and washback	Handbook for the English Classroom: Assessment (Harris, M. & McCann, p) Handbook of Assessment for Language Teachers (Dina Tsagari, et al) Measurement and Assessment in Teaching (Miller et al) Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson) Practical Language Testing (Glenn Fulcher) Language Assessment Principles and Classroom Practices (Douglas Brown) Language Assessment Principles and Classroom Practices (Douglas Brown) Mehrens, W., A. & Lehmann, I., J.)
10	Performance assessment, designing rubrics	Measurement and Assessment in Teaching (Miller et al) Classroom Assessment in Action (Shermis & Vasta) Classroom Assessment Concepts and Applications (Russel & Airasion) Language Testing and Assessment: an advanced resource book. (Glenn Fulcher and Fred Davidson)

APPENDIX E. REFLECTION REPORT TEMPLATE

Name or Nickname:

Date:

Topic of the week:

Through this discussion session I have learnt that

-
-
-

I haven't learnt anything from this reflective reading and discussion session because:

What I found the most effective about this reflective reading and discussion session was:

From this reflective reading and discussion session, I will do

From this reflective reading and discussion session I won't do

In my opinion, this reflective reading and discussion session would have been more effective if

Taking this specific reflective reading and discussion session into account, I do not agree with

APPENDIX F. CLASSROOM OBSERVATION FORM

Announced Observation Policy

Classroom observation will be conducted to observe the participant instructors considering some actions or behaviors stated by themselves in the semi-structure interview. This observation is conducted by the researcher who is one of the instructors at the same institution. A checklist was developed by the researcher in the light of statements given by the participants. I would like to kindly inform that the lesson which will be observed will be recorded through a camera. Researcher as a participant of the lesson won't intervene the flow of the lesson that is it will be a non-participant observation. While the lesson is being recorded, researcher can take some field notes if they are necessary and fill in the checklist.

Procedure

The procedure for the observation will be as follows:

- The instructor to be observed will be informed about the week when the observation is scheduled via email.
- The instructor is expected to conduct an ordinary lesson during observation.
- The observation aims to see the natural flow of the lesson observed That is, there is no need for the instructors to prepare a lesson beforehand. The researcher will try to document whether the instructor as the participant will put what (s)he stated during the interview into practice or not.
- The observer will stay in class until the end of the lesson observed. He will observe the instructors based on the checklist prepared beforehand and he will complete it and he will take field notes.
- If the lesson observed is not within the framework of the plan prepared by the instructor or if the instructor has some questions, at the end of the lesson the instructor will provide with his questions to the instructor.
- In addition to these, after the observation, a follow up interview will be held on a pre-determined day.

Instructor:	Inst.
Observer:	Inst. Yaşar Üstün KAPLAN
Observation Date and Class Hour:	
Research Question:	
Duration of the Observation:	

Before Observation

(1) What are the specific teacher application(s) that you planned to observe for this lesson?

Our research question is aiming at finding out teacher's perception of assessment. Before the observation, we conducted a semi-structured interview and teachers indicated some practices which are related to assessment and are done by her during her lessons.

OBSERVED?	YES	NO	COMMENTS
1.			
2.			
3.			

Post – Observation/ Reflection

1. What do you think about the lesson observed in general?
2. Field-notes taken during observation

APPENDIX I. SEMI-STRUCTURED INTERVIEW QUESTIONS for EXPLORATION of TEACHERS' CURRENT NATURE of COGNITIONS REGARDING LANGUAGE ASSESSMENT LITERACY

1. What do you think assessment is? How would you define assessment? What comes to mind when you think of assessment?
2. How would you define classroom assessment?
3. What techniques do you know about assesment?
4. Which techniques do you use related to assessment?
5. What do you think is the most effective assessment technique? Why?
6. For what purposes do you use the assessment techniques you use?
7. Have there been any changes in the assessment techniques you have adopted since you started teaching? If yes, how? If not, why?
8. When is the best time for assessment in your opinion? Why?
9. What do you think about alternative assessment techniques?
10. What do you think about the assessment results? Are they important to you? Why?
11. For what purposes do you use the assessment results?
12. How do you think an effective assessment should be?
13. What should teachers do for effective assessment?
14. What are your reasons for adopting and using the assessment techniques you use in your classroom practices?
15. Did our country's education policy have any impact on your perception of assessment? If yes, what effect did it have?
16. What kind of assessment does your institution expect from you? So what are you doing?
17. Did the evaluation policy of the institution you work for have an impact on your assessment perceptions? If yes, how?

APPENDIX J. SEMI-STRUCTURED INTERVIEW QUESTIONS for EXPLORATION of TEACHERS' STRENGTHS and WEAKNESSES REGARDING LANGUAGE ASSESSMENT

1. Considering our interview, I learned that you know some basic things about assessment. I saw that you associate the assessment with more classical paper and pencil tests. In fact, assessment is more than just giving a teacher and a test. Do you know this? Do you have an opinion on this?
2. To what extent do you consider yourself competent in choosing and developing an assessment technique that is suitable for the course, useful, easy to apply, adequate and fair? Do you believe that you have enough information on this subject? What do you think are your strengths and weaknesses on this subject?
3. What do you consider when deciding on the assessment technique you will adopt in the lesson?
4. What is self-assessment? Do you know? Have you been exposed to such a practice while you were a student? Do you apply it to your students? How?
5. What is a portfolio assessment? Do you know it? Have you been exposed to such a practice while you were a student? Do you apply it to your students? How?
6. What does measurement error mean to you? Could you please explain it by giving an example?
7. What does validity mean to you?
8. What do valid and invalid assessment data mean to you? Do you think valid and invalid assessment data affect instructional activities and decisions? If so, how do they affect?
9. What sort of assessment method(s) have you used since you started teaching?
10. What sort of assessment method(s) have you decided to use for your present courses?
11. What do you pay attention to in choosing your assessment method(s)? (Ask if not mentioned) Do you pay attention to measurement error and validity?
12. Do you make any research about the method(s) before choosing it/them? If so, what kind of sources do you use?
13. Do you do any research on assessment techniques before choosing them? What do you pay attention to when choosing measurement techniques? Do you pay attention to measurement error and validity?

14. For what purpose(s) do you use the chosen assessment method(s)? Like measuring information, analysing, or making a synthesis.
15. What do you think about the strength(s) and weakness(es) of the assessment method(s) you have chosen?
16. How do you prepare your exam(s) relevant to the assessment method(s) you have chosen? The things paid attention to. / Detailed information about the preparation process.
17. What type(s) of questions do you use in your exams? Why do you use it/them?
18. How do you make your exams valid?
19. Are you preparing a multiple choice test item? How do you adjust the item difficulty value and item discrimination level in these test items? Considering your knowledge on this subject, what are your strengths and weaknesses?
20. What are your strengths and weaknesses in the process of preparing valid and reliable tests?
21. We all prepare tests in our assessment practices. There are different formats of test such as subjective tests and objective tests. What do you consider when preparing the tests? For example, do you consider Bloom's taxonomy when constructing tests?
22. There are standardized tests; they are divided into criterion-referenced and norm-referenced. What do you know about them? If you evaluate your knowledge about this topic, what are your strengths and weaknesses about it?
23. Sometimes you assess students' speaking, sometimes writing, that is, any of their skills. In other words, you assess the performance of students. Do you use rubrics when assessing these performances? If yes, why and how? If not, why?
24. Do you know how to prepare a rubric? So, do you know how to prepare a rubric for performance assessment? How do you ensure the quality of the rubrics?
25. How do you interpret the data you obtained from assessment and evaluation practices?
26. How effective do you see yourself in implementing assessment and evaluation methods, scoring them, and interpreting the data you get from these methods?
27. Do you feel competent in establishing a valid and reliable scoring process?
28. Do you analyse the data you obtain from the assessment and evaluation practices you apply in the classroom? If yes how, if no why? What do reliability and validity mean to you?
29. How do you make your exams reliable?

30. How do you administer your exams? What do you pay attention to in administering your exams? How do you deal with any problem you encounter in administering your exams?
31. How do you score your exams? What do you pay attention to in scoring your exams?
32. How do you grade your exams? What do you pay attention to in grading your exams?
33. Item analysis... Have you heard this term anywhere? What do you know about this subject?
34. What does the consistency of interpreting exams mean to you? Why? (If the concept is known) How do you make your interpretations consistent? (If the concept is known) What do you do in case of inconsistency in interpreting the results?
35. How do you interpret formal and informal students' assessment results? Do you use the results to improve your assessment tool? How? Do you use the results to find out and improve your students' weaknesses? How?
36. Do you take into account the effect(s) of your exams on your students? If yes, how? If no, why?
37. How do you keep your exams and exams' results confidential?
38. What is your attitude toward exam complaint? How do your students make exam complaints?
39. How should assessment and evaluation data be evaluated according to you? What does correct and wrong evaluation mean to you? Please, give examples.
40. Do you use assessment and evaluation data to improve instructional plans to improve your students' learning? If yes, how? If no, why?
41. Do you make any change in your instruction and curriculum? If yes, how? If no, why?
42. What is/are the grading system(s) in your courses? Why do you use this/these system(s)? How have you decided its/their components?
43. If you have not developed any grading system for your courses, what do you think about the grading system(s) given to you? Did you take part in its/their development process? If so, how was/were the grading system(s) developed? Do you use the available grading system without making any change? If not, why? Do you make any change? Please, explain
44. What do you think about the validity of the grades you give to your students by using this/these grading system(s)? How do you explain the grades' validity to your students?
45. Why do you use the grades?

46. How do you announce your exam results to your students and administrators?
47. What do the results mean to your students and administrators according to you? How should assessment results be evaluated?
48. Do you think the chosen assessment method(s) has/have an effect on evaluating the results? Is there any limitation in evaluating the results? If so, what is/are it/they?
49. What do you think about the reflection of evaluating the results?
50. How do you think the misinterpretation and misevaluation of assessment results can be avoided?
51. Do you inform your students and administrators about how to evaluate assessment data?
52. How do you deal with any possible measurement error in communicating your assessment results?
53. What do you think about the ethical, legal and professional behaviours a teacher should follow in assessing and evaluating his students?
54. Do you encounter any problem in terms of your answer to the first question? What is/are the problem(s)? How do you deal with this/these problem(s)?
55. What do you think about ethical issues related to assessment and evaluation? How well do you know about this subject?
56. Do you consider ethical issues in your classroom assessment practices, how?

APPENDIX K. PRE-SEMI-STRUCTURED INTERVIEW IMPLEMENTED FOR EXPLORING TEACHERS' COGNITION DEVELOPMENT CONSIDERING LANGUAGE ASSESSMENT

1. After these sessions, was there any change or improvement on your assessment perception? If yes, how? If not, why?
2. There are several terms related to assessment and evaluation. Like testing, measurement, assessment, evaluation. What do these mean to you?
3. What is formative assessment, what is summative assessment? What is the difference between these two concepts? What can you say about them after the RRD sessions?
4. What is classical - traditional assessment? What is contemporary assessment? Can you define this distinction?
5. What do formal assessment and informal assessment mean to you? Can you give examples of them?
6. What is alternative assessment? What are the techniques for this type of assessment? After the RRD sessions, did you have any progress on this issue? How?
7. After the RRD sessions, do you feel competent in choosing and developing the appropriate assessment technique or tool for your objectives?
8. What do you pay attention to when choosing the assessment technique / tool you will use in the lesson or when developing this technique / assessment tool?
9. Do you know what self-assessment is? If you were to use it in your classroom assessment practices, how would you use it?
10. What is the measurement error according to you? What does it mean to you? (Sources, types of measurement error)
11. What is validity according to you? Have there been any changes or improvements in your knowledge with validity after these sessions? If so, why not?
12. What do valid assessment results mean to you, what do non-valid assessment results mean?
13. In the previous interview, you mentioned the assessment techniques and tools that you have been using since you started teaching. After these sessions, will there be any changes in the assessment techniques and tools and the way they are applied? If so how / if not why?
14. What is portfolio assessment according to you? For what purposes is it used? Have you learned how to apply it?

15. What is your purpose of using the assessment technique you have adopted? In the previous interview, you said "I can determine to what extent the student has learned the subject". Do you still agree, or has there been any change or improvement in your opinion concerning the issue?
16. What are the strengths and weaknesses of the assessment techniques / tools you adopted as a result of the reflective reading and discussion sessions? Considering your readings and discussions, how would you evaluate them? Why?
17. Which technique will you add to your practice as a result of reflective reading and discussion sessions about the assessment, and why?
18. What do you think should be considered while preparing the exam?
19. What kind of questions do you ask in your exams? In other words, what kind of questions do you intend to ask as a result of these sessions, and why?
20. How do you ensure validity in the questions you ask?
21. What do you pay attention to when preparing multiple choice questions? Did the reflective reading and discussion sessions have any effect on you in this regard? If not, why? If so, how?
22. What is a performance assessment? Have your readings on this subject and the discussion you attended afterward caused any development / change related to the subject? If so how, if not why?
23. Do you think you have enough knowledge about rubric now? Do you now feel competent in preparing a rubric or adapting an existing rubric? How?
24. How do you ensure the quality of the rubrics you will use or how will you be sure of the quality of these rubrics after these sessions?
25. For what purposes do you interpret the data you obtained as a result of assessment and evaluation? Or, after the reflective reading and discussion sessions, for what purposes and how should these results be interpreted in your opinion?
26. What does reliability in scoring mean to you? How should reliability and validity be ensured in scoring?
27. What does data analysis mean to you after the RRD sessions you attended?
28. What do you think item analysis is?
29. What is the item difficulty level and item discrimination level?
30. How do you ensure reliability in your exam applications? What are the ways to ensure reliability in administering exams? How will you ensure this after the sessions you attend?

31. How do you think the problems encountered during the exam should be intervened?
Why?
32. What should be considered when scoring? How will you do your scoring from now on?
33. How should grading be done? What should be considered when grading students, why?
34. What does consistency mean to you? How do you think consistency is achieved? What do you think should be done when exam results are inconsistent?
35. How should the impact of exams on students be considered?
36. How should the data obtained as a result of the assessment be evaluated?
37. Can the obtained results cause any changes in the curriculum, instruction? If so how, if not why?
38. How will you be reliable when grading students?
39. How are exam results announced? In the previous interview, you said through the student information system? After these reflective reading and discussion sessions, in what ways and how do you think this sharing should be?
40. What should assessment and evaluation results mean for teachers? What should they mean for students and for administrators?
41. Does the assessment tool or technique used have any effect on the interpretation of the assessment results?
42. How should the reflection of the assessment results obtained be?
43. How can misinterpretation of assessment results be prevented?
44. How can error inclusion in assessment results be avoided?
45. What should be done to behave ethically in assessment and evaluation?
46. As a result of these reflective reading and discussion sessions, do you think you have overcome your lack of ethical issues in assessment and evaluation? If yes how, if no why?
47. In your opinion, what should be the effective and valid assessment and evaluation procedure?
48. Did these reflective reading and discussion sessions, which lasted 10 sessions in general, cause any effect / improvement / or change in your knowledge and cognition level related to assessment and evaluation? If so, how?
49. Do you think these RRD sessions will also affect your classroom assessment practices? If yes, how? If not, why?