



**AN INVESTIGATION INTO THE ASSESSMENT CONCEPTIONS OF
IN-SERVICE ENGLISH LANGUAGE INSTRUCTORS**

İrem Aksu Çekiç

MASTER'S THESIS

DEPARTMENT OF FOREIGN LANGUAGE EDUCATION

ENGLISH LANGUAGE TEACHING PROGRAMME

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

JULY, 2024

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin her hakkı saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren on iki (12) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı: İrem

Soyadı: Aksu Çekiç

Bölümü: İngiliz Dili Eğitimi

İmza:

Teslim Tarihi:

TEZİN

Türkçe Adı: İngilizce Dersi Veren Öğretim Görevlilerinin Ölçme ve Değerlendirme Kavrayışlarına Dair Bir Araştırma

İngilizce Adı: An Investigation into the Assessment Conceptions of In-Service English Language Instructors

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduđumu, yararlandıđım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiđimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduđunu beyan ederim.

Yazar Adı Soyadı: İrem Aksu Çekiç

İmza:

JÜRİ ONAY SAYFASI

İrem Aksu Çekiç tarafından hazırlanan “An Investigation into the Assessment Conceptions of In-Service English Language Instructors” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Yabancı Diller Eğitimi Ana Bilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Müzeyyen Nazlı Güngör

(İngiliz Dili Eğitimi Ana Bilim Dalı, Gazi Üniversitesi)

Başkan: Doç. Dr. Betil Eröz

(İngiliz Dili Eğitimi Ana Bilim Dalı, Orta Doğu Teknik Üniversitesi)

Üye: Prof. Dr. Kemal Sinan Özmen

(İngiliz Dili Eğitimi Ana Bilim Dalı, Gazi Üniversitesi)

Tez Savunma Tarihi: 07.06.2024

Bu tezin Yabancı Diller Eğitimi Ana Bilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban Çetin

Eğitim Bilimleri Enstitü Müdürü

ACKNOWLEDGEMENTS

Writing a thesis is by no means an easy process; however, I have been lucky to be surrounded by many kind and helpful people whose support enabled me to complete this study.

First and foremost, I would love to express my heartfelt gratitude to my advisor, Assoc. Prof. Dr. Müzeyyen Nazlı Güngör, for being such a good mentor and supporting me continuously throughout this journey full of difficulties. She was my light whenever I felt like I was in darkness and lost. Hocam, I cannot thank you enough for all your patience, guidance, and kindness! Secondly, I want to thank my jury members, Prof. Dr. Kemal Sinan Özmen and Assoc. Prof. Dr. Betil Eröz, for their time and contributions.

I also would like to thank Assoc. Prof. Dr. Filiz Kardiye for helping me with the statistical part of the study and answering all my questions kindly whenever I needed.

I would like to extend my gratitude to TÜBİTAK for their financial support within the scope of 2210-A Scholarship Programme. Their support means a lot for novice researchers like me.

The participants of this study deserve a big thanks as well. I cannot express how grateful I am for their time and support. Thank you for sharing all your great opinions with me and making this study possible!

Moreover, I want to thank my friends that I had a chance to meet thanks to Gazi University for all their help and friendship from the beginning of our MA courses. That is why my special thanks go to Sinem, Ebru, Büşra, and Eylül. It has been a pleasure for me to walk on the same path with them hand in hand.

I want to extend my gratitude to my dearest friends. I want to thank Ezgi for lighting my way up by sharing her academic experiences and uplifting words even when she was overseas and Zeynep for being my go-to person whenever I needed support and help from the very first day of my master's studies. Likewise, I want to thank Hazel and Elif for bearing with me whenever

I complain about the challenges I have faced both in this process and in my life for more than ten years. I feel so lucky that we share all the ups and downs of life even when we are far away from each other and thank you for all Zoom studies! Finally, I want to thank all my friends who were not close to me locationally but cheered me up during this challenging process through their calls and messages.

Undoubtedly, I am thankful to my parents Serpil Aksu and Aydın Aksu for their love and support. Knowing that they are with me through thick and thin is a great assurance, thank you! Of course, I also would love to thank my sister Sude Aksu. She is my other half, one of my best friends and supporters. Thanks for all the laughter and happiness you bring to my life under any circumstances. I also would like to express my sincere thanks to the members of my second family, Hatice Çekiç, Yunus Çekiç and Yusuf Ekrem Çekiç, for helping and encouraging me.

My biggest thanks go to my love, best friend, and beloved husband Hüseyin Emre. Your everlasting smile, witty sense of humour, and warm heart have always brought joy to my life and filled my heart with happiness whenever I felt stressed and gloomy. I appreciate your answering me in the kindest way even when I ask, “Am I ever going to be able to finish this thesis?” for the 50th time. Your unwavering belief is what has been pushing me to achieve my goals and I will be forever grateful for that, I love you.

**İNGİLİZCE DERSİ VEREN ÖĞRETİM GÖREVLİLERİNİN ÖLÇME
VE DEĞERLENDİRME KAVRAYIŞLARINA DAİR BİR ARAŞTIRMA**
(Yüksek Lisans Tezi)

İrem Aksu Çekiç

GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Temmuz 2024

ÖZ

Bu çalışmanın temel amacı Ankara/Türkiye’deki üniversitelerin hazırlık okullarında İngilizce dersi veren öğretim görevlerinin ölçme ve değerlendirme kavrayışlarının anlaşılmasıdır. Ayrıca, katılımcıların mezun oldukları bölümler ve deneyim yılları da sorularak bu iki faktöre bağlı olarak ölçme ve değerlendirme kavrayışlarında bir değişiklik olup olmadığı incelenmiştir. Son olarak, katılımcıların ölçme ve değerlendirme amaçlı kullandıkları uygulamalar da sorulmuş olup bu uygulamaların var olan kavrayışlarıyla paralellik gösterip göstermediği incelenmiştir. Çalışmada karma araştırma yöntemi benimsenmiştir. İlk olarak, Brown tarafından geliştirilen (2002, 2004, 2008) TCoA-IIIA isimli ve 27 maddeden oluşan 6’lı Likert ölçeği 101 katılımcıya uygulanmıştır. Ardından, her biri üçer kişiden oluşan üç farklı yarı yapılandırılmış odak grup görüşmesi yapılmıştır. Veri üçgenlemesini sağlamak adına araştırmacı kendini değerlendirme günlüğü tutmuş ve buradan bazı alıntılar dahil ederek okuyucuya süreç adına bilgiler sağlamıştır. Araştırmanın nicel kısmındaki veriler normallik analizinden sonra ANOVA ve Kruskal-Wallis H testi kullanılarak analiz edilmiştir. Çalışmanın nitel kısmı ise içerik analiziyle incelenmiş olup elde edilen kodlar Brown tarafından tanımlanmış ve dört faktörden (gelişim, önemsizlik, öğrenci sorumluluğu, okul sorumluluğu) oluşan modelin altında açıklanmıştır (2002, 2004, 2008). Çalışmanın sonuçları katılımcıların en fazla gelişim faktörüyle hemfikir olduğunu göstermiştir. Bununla birlikte, önemsizlik faktörü ikinci en yüksek değere sahiptir. Bu çalışmadaki en düşük değere sahip faktör ise okul sorumluluğudur. Katılımcıların tüm bu kavrayışlarla ilgili fikirleri odak grup görüşmelerinden desteklenerek açıklanmıştır. Ayrıca, ölçme ve değerlendirmeye dair sahip olunan kavrayışların mezun olunan

bölüm ve deneyim yıllarına göre büyük bir değişiklik göstermediği ortaya çıkmıştır. Ek olarak, katılımcıların çalıştıkları kurumlarda kurum tarafından belirlenen çok çeşitli ölçme ve değerlendirme uygulamaları kullandığı görülmüş, katılımcıların var olan bu uygulamaları daha etkili bir hale getirmek için önerileri olduğu ortaya konmuştur. Son olarak, katılımcıların tercih ettiği ölçme ve değerlendirme uygulamalarıyla sahip oldukları kavrayışlar arasında pozitif bir ilişki olduğunun altı çizilmiştir. Bulgulara ilişkin araştırma ve uygulamaya yönelik çıkarımlar ise son bölümde tartışılmıştır.



Anahtar Kelimeler: Ölçme ve değerlendirme, ölçme ve değerlendirme kavrayışı, öğretim görevlileri, İngilizce dersinde ölçme ve değerlendirme, hazırlık okulları

Sayfa Adedi : xvii + 148

Danışman : Doç. Dr. Müzeyyen Nazlı Güngör

**AN INVESTIGATION INTO THE ASSESSMENT CONCEPTIONS OF
IN-SERVICE ENGLISH LANGUAGE INSTRUCTORS
(M.A Thesis)**

İrem Aksu Çekiç

GAZİ UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

July 2024

ABSTRACT

The main aim of this study is to understand the assessment conceptions of EFL instructors teaching at the preparatory schools of universities in Ankara/Türkiye. Also, participants' undergraduate departments and their years of experience were examined to understand if there was a change in their assessment conceptions based on these two factors. Lastly, the participants were asked to indicate their assessment practices, and whether these practices were in line with their existing conceptions was also investigated. The study adopted a mixed-method research design. Firstly, the 27-item 6-point Likert scale named TCoA-IIIA, developed by Brown (2002, 2004, 2008), was administered to 101 participants. After that, three different semi-structured focus group interviews, each of which included three participants, were conducted. In order to provide triangulation, the researcher kept a self-reflection journal and provided the reader with information about the process by including some entries from the journal. The quantitative part of the study was analysed by using ANOVA and Kruskal-Wallis H test after the normality analysis. The qualitative part of the study was analysed by qualitative content analysis, and the codes were explained under the model defined by Brown (2002, 2004, 2008), which consisted of four factors (i.e. improvement, irrelevance, student accountability, school accountability). The results of the study showed that the participants agreed with the improvement factor the most. Moreover, the irrelevance factor had the second highest value. The factor with the lowest value in this study was school accountability. Participants' opinions on all these conceptions were supported through explanations from the focus group interviews. In addition, it was found that assessment conceptions did not alter significantly based on the department they graduated from and their years of experience. Furthermore, it was understood that the participants used a

wide range of assessment practices determined by their institutions, and they had suggestions to make these practices more effective. Finally, the positive relationship between assessment practices preferred by the participants and their conceptions was underlined. Research implications and practical implications concerning the findings were also discussed in the last chapter.



Key Words : Assessment, assessment conceptions, instructors, English language assessment, preparatory schools

Page Number: xvii + 148

Supervisor : Assoc. Prof. Dr. Müzeyyen Nazlı Güngör

TABLE OF CONTENTS

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU.....	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
JÜRİ ONAY SAYFASI.....	iii
ACKNOWLEDGMENTS.....	iv
ÖZ.....	vi
ABSTRACT.....	viii
TABLE OF CONTENTS.....	x
LIST OF TABLES.....	xv
LIST OF FIGURES.....	xvi
LIST OF ABBREVIATIONS.....	xvii
CHAPTER I.....	1
INTRODUCTION.....	1
1.1 Background of the Study.....	1
1.2 Statement of the Problem.....	3
1.3 Purpose of the Study.....	4
1.4 Significance of the Study.....	5
1.5 Definition of Key Terms.....	5

CHAPTER II.....	7
LITERATURE REVIEW.....	7
2.1 Assessment.....	7
2.1.1 Assessment, Testing, Evaluation and Measurement.....	8
2.1.2 Assessment Methods in Language Classrooms.....	9
2.1.3 Teachers' Role in Assessment.....	12
2.2 Language Assessment in Türkiye.....	13
2.2.1 Assessment Courses in Pre-Service Language Teachers' Education.....	14
2.3 Importance of Teachers' Conceptions.....	15
2.3.1 Assessment Conceptions.....	16
2.3.1.1 <i>Improvement Conception</i>	22
2.3.1.2 <i>School Accountability Conception</i>	23
2.3.1.3 <i>Student Accountability Conception</i>	25
2.3.1.4 <i>Irrelevance Conception</i>	26
2.4 Summary.....	27
CHAPTER III.....	28
METHODOLOGY.....	28
3.1 Research Design.....	28
3.2 Research Questions.....	30
3.3 Research Context.....	31
3.4 Participants/Sampling.....	32
3.5 Data Collection Procedure.....	34
3.6 Data Collection Instruments.....	35
3.6.1 Quantitative Data Collection Instruments.....	35

3.6.1.1 Teachers' Conceptions of Assessment Abridged Scale (TCoA-IIIA)...	36
3.6.2 Qualitative Data Collection Instruments.....	36
3.6.2.1 Focus Group Interviews.....	37
3.6.2.1.1 Pilot Study.....	38
3.7 Researcher's Self-Reflection Journal.....	39
3.8 Data Analysis.....	40
3.8.1 Analysis of Quantitative Data.....	40
3.8.1.1 Validity and Reliability.....	43
3.8.1.1.1 Validity Analysis of the Scale.....	43
3.8.1.1.2 Reliability Analysis of the Scale.....	45
3.8.2 Analysis of Qualitative Data.....	46
3.8.2.1 Credibility and Trustworthiness.....	49
3.9 Limitations of the Study.....	51
CHAPTER IV.....	52
FINDINGS AND DISCUSSION.....	52
4.1 RQ1: What are the assessment conceptions of EFL instructors in terms of four components defined by Brown (2002, 2004, 2008)?.....	52
4.1.1 Improvement Conception.....	53
4.1.1.1 Essentiality.....	53
4.1.1.2 Assistance for Teachers.....	55
4.1.1.3 Guidance.....	56
4.1.1.4 Meticulous Design.....	58
4.1.2 Irrelevance Conception.....	58
4.1.2.1 Exams as Inaccurate Indicators of Student Performance.....	59
4.1.2.2 National Contextual Factors.....	61

4.1.2.3 <i>Teacher Competencies</i>	64
4.1.2.4 <i>Ineffectiveness of Pre-Service Assessment Courses</i>	67
4.1.2.5 <i>Assessment Courses in Shaping PSTs' Mindset</i>	70
4.1.2.6 <i>Standardisation Problems</i>	72
4.1.3 <i>Student Accountability Conception</i>	73
4.1.3.1 <i>Including Peer Assessment and Self-Assessment</i>	74
4.1.3.2 <i>Guide for Students</i>	76
4.1.3.3 <i>Proficiency Exam as a High-Stakes Exam</i>	77
4.1.3.4 <i>High-Stakes Assessment and Student Emotions</i>	78
4.1.4 <i>School Accountability Conception</i>	80
4.1.4.1 <i>No Direct Link Between Institutional/Teacher Success and Student Success</i>	81
4.1.4.2 <i>Solutions to Diminish the Feeling of Incompetence</i>	83
4.1.4.3 <i>Washback Effect</i>	85
4.2 RQ1.1- How do these conceptions vary according to instructors' experience years?.....	88
4.3 RQ1.2- How do the assessment conceptions of the instructors differ based on their educational background?.....	93
4.4 RQ2- What are the self-reported assessment practices that instructors apply in their institutions?.....	99
4.4.1 <i>External Factors Affecting Assessment Practices</i>	106
4.5 RQ3- What is the relationship between EFL instructors' assessment conceptions and the self-reported assessment practices they apply?.....	110
4.6 Summary.....	114
CHAPTER V.....	115
CONCLUSION.....	115
5.1 Summary of the Study.....	115

5.2 Implications.....	116
5.2.1 Practical Implications.....	116
5.2.2 Research Implications.....	118
REFERENCES.....	121
APPENDICES.....	134
Appendix 1. Information & Consent Form.....	135
Appendix 2. Demographic Information & TCoA-III A.....	136
Appendix 3. Responses on Google Forms.....	139
Appendix 4. Focus Group Invitation Form.....	142
Appendix 5. Sample Entries from the Researcher’s Self-Reflection Journal.....	143
Appendix 6. Consent Form for the Focus Group Interview.....	144
Appendix 7. Focus Group Interview Questions.....	145
Appendix 8. Coding Process on MAXQDA.....	146
Appendix 9. Sample Codes and Pre-determined Themes.....	147
Appendix 10. Ethics Committee Approval.....	148

LIST OF TABLES

Table 1. <i>Related Studies in the Literature</i>	18
Table 2. <i>Demographic Information of the Instructors (N=101)</i>	33
Table 3. <i>Focus Group Interview Participants</i>	34
Table 4. <i>Data Collection Process</i>	35
Table 5. <i>Duration of the Interview Sessions</i>	38
Table 6. <i>CFA Results</i>	44
Table 7. <i>Reliability Statistics Results</i>	46
Table 8. <i>Frequency Distributions Categorised by Experience</i>	42
Table 9. <i>Frequency Distributions Categorised by Department</i>	42
Table 10. <i>Descriptive Statistics Results</i>	52
Table 11. <i>Tests of Normality (Shapiro-Wilk Test)</i>	88
Table 12. <i>ANOVA Results</i>	89
Table 13. <i>Tests of Normality (Shapiro-Wilk Test)</i>	93
Table 14. <i>Kruskal-Wallis H Test Results</i>	94

LIST OF FIGURES

<i>Figure 1. Tests, assessment, and teaching.....</i>	<i>9</i>
<i>Figure 2. Qualitative data analysis.....</i>	<i>47</i>
<i>Figure 3. Qualitative data analysis spiral.....</i>	<i>48</i>

LIST OF ABBREVIATIONS

AfL	Assessment for Learning
AoL	Assessment of Learning
AaL	Assessment as Learning
MoNE	Ministry of National Education
PSTs	Pre-Service Teachers
ELTEP	English Language Teaching Programmes
EFL	English as a Foreign Language
ELT	English Language Teaching
ELL	English Language & Literature
CoA-III	Teachers' Conceptions of Assessment Scale
TCoA-IIIA	Teachers' Conceptions of Assessment Abridged Scale
N	Sample Size
CFA	Confirmatory Factor Analysis
ANOVA	Analysis of Variance
M	Mean
I	Interview
ESP	English for Specific Purposes

CHAPTER I

INTRODUCTION

This chapter will provide the framework of the present study, first, by providing background information and the statement of the problem. Then, the purpose of the study and its significance will be discussed. Finally, some essential definitions will be provided.

1.1 Background of the Study

Lambert and Lines (2000) define assessment as a procedure through which teachers collect and interpret data concerning students' responses to tasks in educational contexts. This procedure includes both formal elements, such as tests, and less formal elements, such as evaluating students' sentences while listening to them. This, in turn, helps teachers provide their students with feedback, grade them, or make necessary changes in their teaching (Tosuncuoğlu, 2018). Thus, it is an integral part of the teaching process for many parties, such as policymakers, teachers, and students (Vardar, 2010).

Though assessment and the processes to gather information on students' knowledge and motivation are quite demanding, assessment-related skills, such as an awareness of assessment techniques and methods, are necessary for all teachers to see whether the students have achieved the outcomes that were previously planned and intended (Yetkin, 2017). Also, the idea that teachers' assessment practices affect their students' well-being and learning is underlined by Kyttälä et al. (2022). This leads us to take a deeper look at the assessment practices preferred by teachers as well as their beliefs and conceptions of assessment. Similarly, Zaimoğlu (2013) states that understanding assessment and its meaning is significant as teachers' beliefs on assessment may decide many factors, such as classroom practices, the curriculum to be taught, and what students should achieve. Concerning assessment practices, it can be underlined that a range of different assessment practices can be preferred, and assessment has been described and divided into a variety of categories by many scholars. Summative, formative assessment;

traditional, alternative assessment; or Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL) are some of the example terms included in the assessment literature.

The first categorisation is about summative assessment and formative assessment. The main aim of summative assessment is to give ideas about students' performance and see how much a student knows by evaluating the learning outcomes. This is about providing students with grades and awards or selecting them when necessary. Thus, summative assessment is associated with national exams and standardised exams. On the other hand, formative assessment is concerned with the process of teaching and learning by giving feedback about the course of teaching, which might lead to changes in the teaching procedures if essential. This type of assessment includes asking questions and getting answers in lessons, teachers' observations as well as homework or quizzes (Monteiro, Mata & Santos, 2021). However, some other terms were coined by the Assessment Reform Group (ARG) by changing formative assessment to AfL. Moreover, instead of summative assessment, the term AoL was preferred (Schellekens et al., 2021), which will be comprehensively discussed in the second chapter of the study.

When it comes to the second categorisation, traditional assessment includes some old-fashioned techniques, such as multiple-choice exams and true-false questions, which do not raise higher-order thinking skills, such as analysis and synthesis (Dikli, 2003). On the other hand, McNamara (2001) defines alternative assessment as keeping distance from standard multiple-choice exams and adopting performance-based assessment. In alternative assessment, learners are equipped with higher-level thinking skills, along with problem-solving skills, by focusing on both the process and the product. Moreover, alternative assessment involves a range of methods, such as journals, videotapes, diaries, portfolios, and peer assessments (Çakır, 2013).

In short, teachers prefer to utilise different types of assessment practices in their classes. Their decisions might be influenced by a range of factors, including their beliefs and understanding of assessment. In this vein, to understand how teachers perceive assessment, Brown (2002, 2004, 2008) proposed four different assessment conceptions: improvement, school accountability, student accountability, and irrelevance, which will also be further explained in the second chapter of the study. Brown (2002) also highlights that one of the significant necessities in the professional development of teachers is the comprehension and communication of assessment information. He claims that it is of paramount importance to

understand what teachers believe concerning assessment. With the help of this information, teacher professional development programmes can be designed and applied.

1.2 Statement of the Problem

In Türkiye, the English language starts to be taught in the 2nd grade (MoNE, 2018). This language-teaching process continues for many years. In most universities, foreign language teaching proceeds with preparatory schools and even after the preparatory school year during the following years of students' university education. Kırkgöz (2007) states this by saying that English is the only compulsory language at all levels of education in Türkiye. Based on the need for language teaching, which takes place in every step of the education system, language teachers are employed by many different state and private schools or some other institutions, such as language courses (Yaman & Şahin, 2019). Furthermore, language assessment is inevitable in language teaching. As Tosuncuoğlu (2018) suggests, "Assessment has an important role in education, and it has a critical role in the teaching process" (p. 163). Besides, language teachers are generally responsible and complementary in the assessment process, as stated by Mertler (2003). With this emphasis on the importance of language teaching and assessment and the burden on teachers, a thorough examination of assessment in the teaching and learning context is necessary for many reasons.

Firstly, the language assessment can be problematic for those responsible for implementing the assessment (Mertler, 2003). As mentioned above, teachers are the parties responsible for most of the issues. Therefore, it is necessary to understand how teachers/EFL instructors perceive assessment and the way they put their beliefs into practice since this might pave the way for teacher development programmes when necessary.

Secondly, Harris and Brown (2009) point out that the conceptions that teachers have affect their assessment practices, and these assessment conceptions might be shaped by the courses they took during their pre-service education. With this study, the question of whether the curriculum and courses of English Language Teaching programmes (ELTEP) prepare pre-service teachers (PSTs) for their in-service experiences in a sufficient way in terms of assessment will be analysed because most of the time, both during their studies as PSTs or as in-service teachers after their graduation, teachers do not feel adequate in terms of assessment skills. Additionally, the influences of these courses taken by the instructors as undergraduate students on their assessment conceptions will be understood with this study since it might be necessary to change

the number or the content of those assessment courses to prepare English Language Teaching (ELT) students for their in-service teaching experiences. For instance, the researcher of this study is familiar with one of the Foreign Language Education departments of one of the well-known universities in Ankara/Türkiye, which included only one assessment course in the curriculum implemented before 2020, which might not be enough to equip the students with the necessary knowledge about assessment practices as well as effectively shape their assessment conceptions.

Lastly, teachers, most of the time, are held responsible for the implementation of assessment practices that were previously defined by the institutions themselves, which leaves little room and flexibility for teachers, and this may make teachers act opposite of what they already believe. Therefore, learning teachers' opinions and perceptions on the already-existing practices help reconsider the effectiveness of assessment techniques and procedures being conducted at their institutions. With feedback and an in-depth understanding of assessment conceptions of teachers, institutions can enrich and shape their future assessment practices.

1.3 Purpose of the Study

Assessment is an integral part of teaching and learning practices for all levels of students, as previously mentioned. Being exposed to assessment throughout their academic lives, students have beliefs about assessment. However, students are not the only ones holding beliefs/assumptions concerning assessment. Teachers also have certain conception/s regarding assessment that they reflect on their assessment practices (Brown & Harris, 2012). Assessment conceptions of teachers might be affected by different factors.

To begin with, teachers' beliefs can be affected by their own experiences when they were students. In addition, courses offered to them when they were PSTs might pave the way for them to have specific ideas on assessment. Nevertheless, these conceptions can also alter over time as more experience is gained as in-service teachers (Xu & He, 2019). Lastly, it should also be underlined that from time to time, institutions or policies of countries where teachers work may oblige them to follow the assessment practices previously decided and defined by them. The culture and the educational system may shape their conceptions (Kyttälä et al., 2022) or teachers may apply what is asked of them even though they might not favour those practices.

Concerning all those, the main aim of the study is to find out the assessment conceptions of English as a foreign language (EFL) instructors working at preparatory schools of six different

universities in Ankara in the 2022-2023 and 2023-2024 academic years based on Brown's four-facet assessment conception model (2002, 2004, 2008) by relating them with their:

1. years of experience,
2. educational background,
3. classroom practices.

1.4 Significance of the Study

This study is significant in many regards. Firstly, teachers' conceptions lead them to shape their assessment practices in their classes (Harris & Brown, 2009). Thus, the information related to assessment conceptions and assessment practices of EFL instructors working at the tertiary level will be explored and interpreted in this study.

Secondly, even though there are studies concerning assessment conceptions of teachers in different contexts with teachers teaching different subjects (Lutovac & Flores, 2022; Monteiro et al., 2021), the number of research studies investigating EFL instructors' assessment conceptions and assessment practices at the same time in the Turkish context is limited. This makes the study an important one.

Last but not least, having a comprehensive idea of what teachers prefer and do not prefer regarding assessment practices can lead the institutions to make some changes in their programmes, if necessary, to increase the effectiveness of the assessment processes. In this vein, Brown (2002) clarified that in-service training can also be prepared based on teachers' conceptions of assessment and their reflections on their assessment. Kyttälä et al. (2022) support this by stating that understanding assessment conceptions and reflecting on them is required, along with practical and theoretical knowledge to support change in conceptions, which is related to teacher education.

1.5 Definition of Key Terms

Conceptions: Conceptions, in the context of this study, involve teachers' opinions, attitudes, and beliefs (Thompson, as cited in Brown, Hui, Yu & Kennedy, 2011).

Assessment: A notion in which various techniques are used to interpret information about students' performances (Brown, 2002).

Evaluation: Evaluation is deciding the efficiency of students and services with the help of the data gathered by assessment, and while assessment is defined as a method, evaluation is defined as a process/activity (Gottlieb, 2016; Tosuncuoğlu, 2018).

Measurement: “Measurement is situating data from an assessment in a quantitative framework, to characterize the evidence the observations provide for the interpretations and inferences the assessment is meant to support” (Mislevy, 2017, p. 11).

Testing: Gathering data from test-takers fairly and scoring this data by following certain rules to make deductions about an ability/performance and understand the needs (Brown, 2019).

English Language Instructor: The staff who teaches English at the school of foreign languages/preparatory schools of universities. In this study, the words “*teacher*” and “*instructor*” are used interchangeably.

CHAPTER II

LITERATURE REVIEW

As this study aims to understand EFL instructors' assessment conceptions, some background information will be provided concerning assessment, assessment-related notions and various assessment practices that can be applied in language classrooms. Secondly, the importance of teachers in assessment, assessment in the Turkish context, and the significance of teachers' assessment conceptions will be underlined by explaining four different assessment conceptions covered under the scope of this study. This will help readers to have a frame in their minds to comprehend the current study better. Furthermore, different research studies in the literature from Türkiye and abroad concerning assessment and assessment conceptions will be exemplified to foster this understanding.

2.1 Assessment

Assessment, which plays a crucial role in teaching and learning, has been defined and explained frequently in the literature with different words (Azis, 2012). Gülbahar and Büyüköztürk (2008) define assessment as an umbrella term that includes a lot of different practices and activities. Furthermore, it is defined as “the process of collecting information about a student to aid in decision making about the progress and language development of the student” by Cheng, Rogers, and Hu (2004, p. 363). Purpura (2016) lists some features of assessment. First, it is stated that assessment is designed to understand information based on verifiable content. Second, even though the purpose can occasionally be seen implicitly or explicitly, assessments are based on a goal. Lastly, assessments can take many forms, from controlled to less controlled practices (Purpura, 2016). In addition, the importance of assessment in understanding and evaluating existing language teaching practices is unquestionable, as Ali (2011) indicates. He argues that teachers use assessment methods to understand their students' current level and

decide the areas to be improved for further learning, which is essential for both students and teachers.

Similarly, Yetkin and Özer (2020) underline that assessment not only shows the amount of learning but also leads to improvements in teaching to achieve the pre-decided outcomes. Finally, the importance of assessment is underlined in a relatively recent research study by Kyttälä et al. (2022). They state that the interaction between learning and assessment has been increasingly observed as time passes, making assessment integral for all learning situations.

2.1.1 Assessment, Testing, Evaluation and Measurement

Assessment is a widely studied topic in education, and it has been described in the literature many times by various scholars. Nevertheless, there is a range of other terms to be considered and defined linked to assessment since these terms are generally used in a similar sense or interchangeably, which might confuse readers of this study or even teachers themselves, as illustrated by Pastore and Mincu (2024). Therefore, the terms assessment, testing, measurement, and evaluation will be explained as the distinction between those should be clarified to grasp what assessment means in a better way.

Firstly, H. D. Brown (2004) points out the difference between assessment and testing. He says that while tests are pre-designed administrative practices by which students know they are evaluated, assessment covers a broader concept. In other words, teachers assess their students continuously with their answers and performances during lessons and while checking their homework. This can be a deliberate or a subconscious assessment. Similarly, in Nasab (2015), assessment is explained as an informal process through which information related to students' knowledge is gathered with the help of various techniques in various contexts. This distinction between testing and assessment is also illustrated with a scheme below:

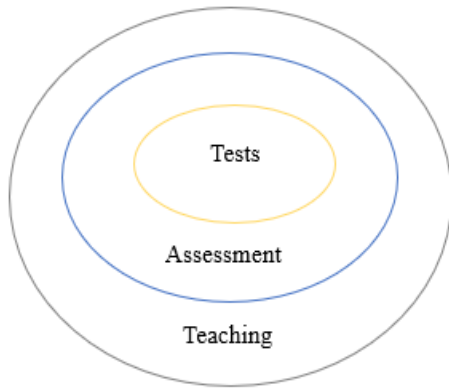


Figure 1. Tests, assessment, and teaching. Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson Education.

Secondly, evaluation and assessment are defined differently in the assessment literature. Huitt, Hummel, and Kaeck (2001) describe evaluation with the following words: “Evaluation refers to the comparison of data to a standard for the purpose of judging worth or quality” (2001, p. 1), while assessment is about collecting data to describe or understand an issue in a better way (Huitt, 2007).

Lastly, the distinction between measurement and assessment can be clarified. Kizlik (2019) defines measurement as a process to understand the characteristics of a physical object. S/he also adds that there are some specific instruments in measurement, such as scales. In this regard, Yüce (2015) indicates that measurement is exact, while assessment is not.

2.1.2 Assessment Methods in Language Classrooms

Assessment is considered one of the crucial elements of foreign language classes, and different approaches and angles should be considered to improve learners’ skills in a foreign language (Çakır, 2013). Different perspectives and approaches in assessment may lead to different results. Black and Wiliam (1998) support this idea as they highlight the significance of assessment practices. They indicate that with the help of assessment practices, student learning can either be improved or hampered. Furthermore, Khodashenas, Khodabakhshzadeh, Baghaei, and Motallebzadeh (2022) argue that understanding and gaining an awareness of assessment practices and their meticulous application substantially impact language learning programmes and there are different classification systems concerning assessment practices (Zaimoğlu, 2013).

To begin with, Harlen and James (1997) discuss summative and formative assessments. These two types of assessments have different natures and practices. Summative assessment is generally concerned with traditional ways of assessment including using assessment results for high-stakes purposes (Shute & Kim, 2014). Hence, scoring and grading are included in summative assessments. Final exams and university entrance exams can be given as examples of this type. No other exams take place if students succeed in summative assessment, and it aims to report how much a student has learned to parents, other teachers, and people in charge, as well as students themselves (Dixson & Worrell, 2016; Zaimoğlu, 2013).

On the other hand, formative assessment is conducted to see where learners are in terms of a specific skill, and learners' progress has significance as well as their efforts in the process (Harlen & James, 1997). Stiggins (2005) states that formative assessment, also called AfL, is used in learning to promote it. The main aim of formative assessment is not to grade students or judge them. Therefore, this assessment type includes self-assessment, peer assessment, presentations, and portfolios (Al-Nouh, Taqi & Abdul-Kareem, 2014). Greenstein (2010) indicates that with the help of formative assessment, teachers can comprehend the emotional needs of their students. Moreover, it enables teachers to observe differences between students throughout the teaching process and helps students gain self-confidence with social and emotional support.

Concerning this, Black and Wiliam (2018) underline that assessment with the purpose of improving classroom practice should be focused on the process of learning and assessment by integrating all the aspects of assessment. Similarly, Gebril (2016) points out that in-service teachers prefer a combination of summative and formative assessments.

Another similar form of classification is traditional and alternative assessments. The traditional assessment form is associated with techniques, such as multiple-choice exercises, true-false activities, and others. In these techniques, some points still need to be added, such as detailed information on the process and the product of learning, as Nasab (2015) discussed. Besides, these activities do not resemble real-life situations in which language is used, which might lead them to be labelled as inauthentic (Nasab, 2015). Another drawback of traditional assessment techniques is that utilising only traditional assessment methods, such as exams in pen and paper form, may not stimulate students' motivation. Furthermore, test results may not accurately indicate students' language skills. For instance, there might be some cases in which students

can get high scores in pen-and-paper format exams, but they might have problems while using the language in real life (NCLRC, 2004, as cited in Phongsirikul, 2018).

On the other hand, Norris, Brown, Hudson, and Yoshioka (1998) state that, unlike traditional assessment, alternative assessment addresses students' problem-solving and higher-level thinking skills by including real-world contexts. The emphasis is both on the process and the product. Dikli (2003) explains different approaches to alternative assessment, such as authentic assessment and performance-based assessment. While authenticity is related to including real-world issues in classrooms to connect real-life experiences to students' lives, the performance-based part is related to students' creation of continuous, apparent responses (Elliott, 1995). However, the terms performance-based assessment, authentic assessment, and alternative assessment are used interchangeably by scholars, as Dikli (2003) suggests. Nasab (2015) highlights two other elements of the alternative assessment: portfolios and projects. Firstly, with the help of portfolios, students' efforts throughout the process and their improvement can be examined since portfolios are helpful in demonstrating students' efforts and success through a collection of students' work. Thus, portfolios encourage student participation and reflection (Delett, Barnhardt & Kevorkian, 2008). Additionally, projects are goal-oriented tasks, and they are advantageous since they make students have plans and use their skills to solve problems or achieve a specific goal. Although these assessment methods are beneficial in many ways, they require more commitment and responsibility from teachers and students because these forms of assessments take more time and are cumulative (Nasab, 2015).

Moreover, some other terms were defined in the assessment literature. These definitions were coined by the Assessment Reform Group (ARG) as alternatives for summative and formative assessment. Their definitions include AaL, AfL, and AoL. To begin with, AfL is described by Klenowski (2009) with these words: "AfL is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration, and observation in ways that enhance ongoing learning" (p. 2). Lee (2011) states that learning is the primary goal in AfL, and assessment is a tool for achieving that goal. Besides, s/he underlines the positive effects of AfL on students' learning, such as higher motivation and performance. Lee (2011) found the positive impact of AfL in writing for both teachers and students. Secondly, though it is sometimes thought of as the sub-branch or the last step of AfL by some scholars, Lam (2015) defines AaL by highlighting that it is about promoting learners' independence by engaging them in assessment-related activities and improving the learning process accordingly. Therefore, it can be briefly described as including students in the learning process, which can

be achieved through self-assessment or autonomous learning. Thus, students have a crucial role in learning and assessment in this view (Schellekens et al., 2021). Finally, AoL is similar to summative assessment. In other words, the aim is to report students' performance to parents or the people in charge, as Lee (2011) indicates. Therefore, it is apparent that AoL is used by administrators for certification or to conclude students' success as well as a medium that shows the success of a school. Accordingly, it is inferred that AoL takes place as exams, whereas AfL can occur in various forms, such as observation and classroom interaction.

2.1.3 Teachers' Role in Assessment

Assessment includes many parties, such as parents, policymakers, and teachers. Despite this, it is generally thought that it is the classroom practitioners' responsibility to conduct assessment practices, which take around 40% of the classroom time, and assessment practices are attached to the work provided by the teacher (Stiggins, 1988). Moreover, some other studies in the literature support this view by underlining teachers' dual role as teachers and assessors and indicating that they are the first to interpret the assessment results regardless of the assessment practices used (Azis, 2012; Levi & Inbar-Lourie, 2020). Hence, teachers are expected to both focus on classroom assessment and be familiar with high-stakes assessment (Levi & Inbar-Lourie, 2020). Regarding this, assessment literacy should also be defined with a highlight on the paramount importance of teachers. Popham (2011) underscores the importance of assessment literacy, which can be briefly described as teachers' knowledge and understanding of concepts regarding assessment, which probably have an impact on their decisions. Moreover, assessment literacy is useful in helping teachers set priorities, give feedback, and use teaching methods effectively (Pastore & Mincu, 2024).

Scholars emphasise teachers' knowledge and skills since teachers are influential decision-makers in the classroom regarding assessment and many other things. Concerning assessment, they decide many things, from marking quiz papers to determining whether a student will pass or fail, even though the significance of teachers' decisions can vary depending on the context. By considering this, it is essential to understand teachers' assessment conceptions and the way they put those conceptions into practice, along with their needs because if teachers do not hold positive beliefs or attitudes towards assessment, or if they are not assessment literate, some negative consequences will be seen in terms of the quality of assessment outcomes. In the same vein, Popham (2009) argues that as the significance that teachers attach to assessment increases,

classroom assessment practices will become more meaningful by incorporating various techniques. Finally, it should also be noted that teachers need to be prepared to use alternative assessment methods successfully as there is a shift from traditional assessment to alternative assessment. Highlighting teachers as assessors leads us to examine the assessment courses provided to teachers to shape their understanding and conceptions of assessment and their knowledge in the related field. In this vein, the assessment context in Türkiye will be clarified first, and language courses provided to PSTs in Türkiye will be explained.

2.2 Language Assessment in Türkiye

The National English curriculum in Türkiye is centred around CEFR and requires a learning environment in which language is taught in a meaningful context through production and communication. Language functions and real-life skills are tackled together. Similarly, language assessment in the curriculum has a parallelism with the teaching part (Çimen, 2022). In other words, language assessment also encourages assessing four skills with the help of a variety of assessment tools including alternative assessment and traditional assessment methods, as argued by Çimen (2022). Similarly, it is stated by Kitchen, Bethell, Fordham, Henderson, and Li (2019) that Turkey underwent a transformation in education by adopting a student-centred approach in 2018, which would also be reflected in assessment methods. This means that the focus would be less on grades and exams, and students' development and progress would be the focal point. However, it is also indicated by Kitchen et al. (2019) that teachers acknowledged using assessment methods that are more practical, such as multiple-choice questions even though they are encouraged to use performance-based assessment. Their lack of knowledge in using a variety of assessment tools is also underlined.

When it comes to the tertiary level, English is used as a tool for communication in most universities in Türkiye. With the increasing number of universities using English as a medium of instruction, a one-year preparatory school programme is offered to students, whose language proficiency levels are weak, to enable them to express themselves both in academic and general contexts (Bayram & Canaran, 2019; British Council & TEPAV, 2015). Concerning assessment, it should be noted that preparatory schools also relate their materials, objectives and level descriptions to CEFR. Nevertheless, the report of the British Council and TEPAV (2015) points to some concerns about assessment that must be addressed in preparatory schools. First, the fact that preparatory schools fail to make their students reach the intended level is indicated in the

report. In other words, students are said to reach B1 level by their teachers, even though the aim is to make them reach B2 level in the first place. Besides, the quality and the coverage of proficiency exams prepared by institutions are some other concerns regarding assessment in preparatory schools, such as the negative washback effect and the policies of schools that teachers have to follow are also pointed out, similar to the findings of this study (See Chapter IV).

2.2.1 Assessment Courses in Pre-Service Language Teachers' Education

Language teaching has been gaining more and more importance in our century, and based on this demand, more English language teachers are employed (Ulum, 2015). In Türkiye, among different job opportunities, English language teachers might prefer working as teachers at the Ministry of National Education (MoNE) or working at universities as language instructors to teach English courses. The success of the education those teachers provide is, in turn, parallel to the quality of teachers (Lutovac & Flores, 2022; Ulum, 2015). To have quality teachers, on the other hand, it is essential to give them decent pre-service education by equipping them with adequate thinking and planning skills. Moreover, they should be able to analyse and evaluate the consequences of their lessons.

Therefore, with teachers playing a significant role in assessment, considering their educational background is crucial since it might lead them to adopt assessment-related beliefs and practices accordingly. This idea is supported by Kyttälä et al. (2022) as they suggest that teachers' conceptions of assessment and assessment skills are constructed when they were students in teacher education programmes, even though these conceptions might change in time with more experiences. Therefore, the quality of pre-service assessment courses is becoming more and more important (Güngör & Güngör, 2024). For that reason, the educational background of foreign language teachers is a subject of discussion.

In Türkiye, after the university entrance exam, PSTs are subjected to a standard four-year ELTEP following a nationwide mandatory curriculum prepared by the Council of Higher Education unless they study in language preparatory schools for an extra year (Hatipoğlu & Erçetin, 2014; Ulum, 2015). It is stated by Hatipoğlu (2015) that the curriculum of the ELTEP includes only one course on language testing and evaluation, leaving lecturers with doubts about the content of this course, along with the time to be spent on the topics. Moreover, including

only one course on assessment is not sufficient to provide PSTs with enough knowledge concerning assessment since limited time leads lecturers to cover the necessary topics in a superficial way (Hatipoğlu & Erçetin, 2014). Comparable to the Turkish context, some studies claim that teachers feel insufficient when it comes to assessment, which might be associated with the education they receive as PSTs (DeLuca, 2012; Ogan-Bekiroglu & Suzuk, 2014), even though assessment is one of the touchstones of teacher education programmes, as stated by many (Kyttälä et al., 2022; Yetkin, 2017). Pastore (2023b) supports this by indicating that assessment literacy has been touted as one of the crucial points of education; however, teachers do not consider themselves confident, and their assessment literacy levels are not high. Besides, it is not possible for pre-service and in-service teachers to put their literacy into practice due to the limited time, syllabus, and contextual restraints (Güngör & Güngör, 2024; Ogan-Bekiroglu & Suzuk, 2014). Similarly, Mertler (2003) points out that teachers feel inadequate in their assessment-related competencies because of their insufficient education as PSTs. Therefore, there is a need to improve teacher education programmes. Additionally, it is stated that the practicum component, which could be an essential component while shaping PSTs' conceptions of assessment, is neglected in the education of PSTs both by scholars from Türkiye and other countries (Güngör & Güngör, 2024; Xu & He, 2019). Lastly, Hung and Wu (2023) also highlight that teachers do not always have sufficient education, and they indicate the insufficient training of teachers both in knowledge and practical terms, which leads to dissatisfaction concerning assessment practices (Khodashenas et al., 2022). Based on all these findings and the significance of assessment, it is indisputable to put some effort into understanding assessment and assessment courses. Likewise, this study will tackle this issue by trying to understand how assessment courses shape the assessment conceptions of EFL instructors, which will be thoroughly explained in the following chapters.

2.3 Importance of Teachers' Conceptions

In general terms, concepts are described as units in which entities are classified based on similar features, which are decided in time via life experiences. These entities differ from others and are collected in the brain as a thinking style. Concepts are needed for people to comprehend the world, and these concepts create the basis for understanding someone's environment, as Meral (2022) indicates. It can be inferred from the same study that mental structures created in the mind concerning a concept are named conceptions (Meral, 2022).

In teaching contexts, Brown (2004) describes conceptions as mental frameworks that teachers possess to interpret and evaluate their teaching environment, which cover beliefs. Investigation of teachers' conceptions is a valuable field and provides an understanding of teachers' decisions, practices, and pedagogy, as Lutovac and Flores (2022) express. Similarly, Barnes, Fives, and Dacey (2017) state that teachers' actions are based on their beliefs or conceptions.

There are different statements concerning teachers' conceptions or beliefs. Firstly, for Clark (1986), teachers' conceptions are imperfect and inconsistent since their experiences may put them in a dilemma without solutions (as cited in Brown, 2008). On the other hand, Levin (2015) indicates that teachers' beliefs and their knowledge are strongly connected. Also, these beliefs can guide them to take their future steps (Fives & Buehl, 2014). Thus, teachers' beliefs are essential in their teaching and students' performance (Comforto, 2022).

In short, teachers' conceptions are worth considering in the scope of this study since teachers have some specific beliefs and conceptions concerning language assessment, which give meaning to their teaching and explain their behaviours (Betoret & Gómez-Artiga, 2004).

2.3.1 Assessment Conceptions

Assessment conceptions are the beliefs and understandings of teachers about assessment (Lutovac & Flores, 2022). As previously mentioned, teachers' conceptions and beliefs about assessment are related to the assessment practices they choose to apply. Likewise, Pastore (2023a) expresses that studies on assessment conceptions paved the way to comprehend teachers' practices.

There are many factors that may affect teachers' conceptions regarding assessment. Firstly, external factors such as the culture or the educational system the teachers are in can influence their conceptions, as Lutovac and Flores (2022) indicate. Because of these reasons, students encounter many different approaches throughout their education (Comforto, 2022). Secondly, it is worth noting that gaining professional and personal experiences, on the other hand, may cause in-service teachers to experience a change in their assessment conceptions as time goes by (Xu & He, 2019). Lastly, it can be added that teachers' experiences as students might affect the assessment conceptions they adopt in their professional lives (Kyttälä et al., 2022). It should also be noted that teachers' conceptions can be more than one, and they can even own conflicting conceptions concurrently.

There are some studies in the literature concerning teachers' assessment conceptions. For instance, in the study by Betoret and Gómez-Artiga (2004), assessment conceptions of teachers were illustrated within four titles: student-centred vs. teacher-centred and process-centred vs. product-centred. (as cited in Pastore, 2023a). Furthermore, some studies investigating the differences in conceptions of pre-service and in-service teachers can be found in the literature, as suggested by Pastore and Mincu (2024). Some additional assessment conceptions studies conducted in the literature have been examined and demonstrated in descending order of their dates in Table 1.



Table 1

Related Studies in the Literature

Author/s & Year	Aim of the Study	Participants	Data Collection Tools	Data Analysis	Findings & Results
Kyttälä et al. (2022)	Finding out the conceptions of assessment of the Finnish teachers who are specialised as subject teachers, classroom teachers, and special needs teachers.	287 participants who were PSTs participated in the study to represent three Finnish universities.	The participants had a web-based 20-item questionnaire in the study including ten items from CoA-III designed by Brown. The last ten items were developed for the Finnish context to comprehend the national issues for assessment. The last form included three categories: 1- Assessment of learning, 2- Assessment for teaching and learning, 3- Assessment as harmful.	A variety of techniques were used for analysis. Firstly, the researchers got help from descriptive statistics and Confirmatory Factor Analysis (CFA). Moreover, MANOVA, ANOVA, and cluster analysis were applied by using K-means method to answer different research questions in the study.	It was stated that the existing conceptions both within the group and between groups were varied. In other words, the conceptions ranged from assessment-oriented and formatively oriented to summatively oriented. Anti assessment-oriented conception was also seen among the participants.
Lutovac and Flores (2022)	Pre-service subject teachers were investigated by using their narratives for student failure.	The participants consist of 79 Finnish pre-service subject teachers.	To analyse the data, teachers' narratives of students' failure were collected in the scope of a course in a teacher education programme.	Content analysis was applied to analyse the data.	Four main conceptions were revealed as a result: 1- Assessment were related to feedback as well as reflection, 2- Assessment should show the learning of the students, and it should be personal, 3- Students' efforts should be taken into account, 4- Assessment was not a way to comprehend students' success and failure.

Table 1

Related Studies in the Literature (Continued)

Author/s & Year	Aim of the Study	Participants	Data Collection Tools	Data Analysis	Findings & Results
Monteiro et al. (2021)	The aim is to explore assessment conceptions of Portuguese teachers and students as well as teachers' practices in mathematics and the relationship between those.	Five teachers from four different schools, which were decided based on purposive sampling, participated in the study. Moreover, the student part of the study consisted of 82 pupils.	Researchers carried out individual interviews with the participant teachers, and they included focus group interviews with the students at the end of the year. Besides, students' documents (worksheets/tests) were used because the researchers wanted to see the assessment practices used by the teachers.	After transcribing all the interviews verbatim, content analysis was applied by two of the authors. Moreover, 24 lessons and 860 documents were analysed. As indicated, an observation grid was utilised to analyse qualitative data. Moreover, the researchers got help from descriptive statistics to differentiate summative and formative assessment techniques.	The results indicated four categories ranging from "focused on accountability" to "focused on learning." Assessment conceptions of teachers and students demonstrated variation, and there was no consistency between teachers and their students. (Apart from one of the teachers.) It was revealed that teachers' main belief was that assessment improved learning and teaching (AfL), whereas students' assessment conception was at the AoL pole by underlining summative assessment. Also, there was no consistency between teachers' assessment conceptions and their assessment practices, while there was some consistency between teachers' assessment practices and students' conceptions.

Table 1

Related Studies in the Literature (Continued)

Author/s & Year	Aim of the Study	Participants	Data Collection Tools	Data Analysis	Findings & Results
Xu and He (2019)	The study aims to see the change in PSTs' assessment conceptions in practicum and the reasons for this change.	The study included three participants from the same practicum team in China. (Due to time limitations, these three people were chosen among six participants.)	Firstly, a focus group interview was conducted to see the initial conceptions of assessment of the participants. After their practicum, the researchers carried out individual interviews to see the changes in their conceptions of assessment and the causes of the change, if there were any. Moreover, the researchers got help from classroom observations and lesson plans.	To analyse the data, within-case and cross-case analyses were used. Concerning within-case analysis, the cases were evaluated individually, and the focus was on their conceptions before and following the practicum and the effective elements. This was followed by the analysis of supplementary data: 1- Watching videos and summarizing classroom activities, 2- Analysis of the summaries and extracting assessment activities, (e.g., feedback delivery) 3- Open coding, 4- To group the codes into categories, axial coding was used. 5- Lastly, this data was used to compare the lesson plans of the participants to see if there was a difference between the plans and the actual lesson practices regarding assessment.	A noticeable change was observed in the practicum of these three teachers. The change in their assessment conceptions can be listed as follows: 1- Comprehensive perception of purposes of assessment, 2- Having realistic perceptions concerning assessment criteria, 3- More understanding in relation to feedback and strategies and their power, 4- More attention to fairness in classroom assessment, involving and engaging students in the assessment process. Also, changes in teachers' conceptions of assessment were different in these cases.

Table 1

Related Studies in the Literature (Continued)

Author/s & Year	Aim of the Study	Participants	Data Collection Tools	Data Analysis	Findings & Results
Yüce (2015)	The thesis aims to examine the PSTs' assessment conceptions and their intended assessment practices.	133 pre-service English language teachers were the participants in the study.	The data were gathered by using the shorter version of the Teacher Conceptions of Assessment Scale (TCoA-III A).	To answer the research questions, descriptive statistics was used by calculating frequencies and percentages. Also, the Pearson Correlation was applied.	It was indicated that some of the PSTs thought that assessment was related to improvement, while most of them found it irrelevant. Also, PSTs in the same study stated that they planned to use student written work (79%) as one of their assessment practices. This was followed by oral question and answer (72%), planned observation (72%), and others.
Remesal (2011)	The aim is to understand the conceptions of in-service teachers about the functions of assessment on four approaches: 1- Assessment influences on teaching, 2- Assessment influences on learning, 3- Assessment influences on accreditation of learning, 4- Assessment influences on teachers' accountability.	50 primary and secondary school teachers participated in the study.	In the data collection process, 50 teachers first underwent a semi-structured interview. Later, the same teachers were interviewed by using an event-recall technique.	All the data were transcribed to be analysed with the help of qualitative content analysis by following a recursive process in which the categories were defined gradually.	The four items mentioned in the "Aim of the Study" column could not be separated as they were intertwined. Teachers also stated contradictory opinions on the effects of assessment on teaching/learning. Also, it was stated that those conceptions might have a relationship with the existing education system and the demands of this system since the beliefs of secondary school and primary school teachers showed differences.

In addition to the studies presented above, Brown (2002, 2004, 2008) proposed four interrelated assessment conceptions and developed a scale to understand teachers' conceptions of assessment (TCoA-III). The shorter version (TCoA-III Abridged Scale or TCoA-IIIA) was used in this study. According to Brown (2008), the conceptions that teachers hold can be divided into four categories:

1. Improvement
2. School accountability
3. Student accountability
4. Irrelevance

Based on this four-facet model, some teachers who have the improvement conception believe that the quality of teaching can be improved with the help of assessment. Teachers may have school accountability conception and believe that assessment results help decide the quality of the schools. Furthermore, some teachers may think that students can be graded based on a criterion to see whether they have achieved the required standards. Nevertheless, assessment can also be perceived as something irrelevant by teachers, which might be because teachers ignore assessment results (Yüce, 2015). These four conceptions are clarified in detail below since they form the basis of this study.

2.3.1.1 Improvement Conception

As Brown (2002) describes, “The major premise of this conception is that assessment informs the improvement of students' own learning and improves the quality of teaching” (p. 27). In other words, it helps get information on students' learning and problems, and it is a way of improving teaching by making necessary changes in the teaching process to facilitate student learning. Assessment should be reliable and valid in representing students' performances (Brown, 2008). Assessment improves learning and teaching, provides feedback, and should be considered positive and constructive (Monteiro et al., 2021).

Assessment comprises teaching, learning, and planning, even though these steps occur at different times (Brown, 2008). The ones with improvement conception may think that teachers have many responsibilities during these phases, which is a requirement for teachers. One of the responsibilities of teachers is to include different assessment techniques in their teaching. Therefore, both summative and formative assessments can be used to improve the quality of

teaching and evaluate students' learning. Various assessment techniques should be used to comprehensively explain what students have learned (Brown, 2008; Moiiinvaziri, 2015). Thus, combining teachers' judgements about students' performance, which is an informal assessment, and formal assessment practices serve the primary purpose of assessment: improving student learning (Brown, 2004). The sole usage of multiple-choice tests and lower thinking skills are considered a point of rejection, whereas higher thinking skills, such as analysis and synthesis, are inevitable elements. Moreover, including students in the learning process is crucial to developing the quality of education (Brown, 2008). Despite being significant, the problem concerning improvement conception is the burden it places on teachers. In other words, teachers' identification of students' performance by including higher-order skills and being actively involved in assessing students' learning process with the help of various methods is necessary.

Another vital point alongside the involvement of higher-order skills is the validity of the assessment results. Messick (1989) defines validity as having parallelism between the curriculum and assessment methods in that respect. Assessment results should also be correct, which should pave the way for the accurate interpretation of the results to pursue effective methods while taking future steps (as cited in Brown, 2008). Validity is ensured when subjectivity in the grading stage is avoided, and consistent assessment materials are used.

In short, the conception of improvement is situated on the thought that assessment is essential to improving learning and teaching through the use of valid assessment techniques. These techniques must elicit students' higher-order skills and should vary to give teachers the best idea of their students' learning.

2.3.1.2 School Accountability Conception

The word accountability was defined by Wiliam (2010) as being responsible, leading to expectations from a person or an institution. Thus, schools are held accountable to many parties, such as parents and students, in educational contexts. The primary concern in this conception is assessment results, as explained by Brown (2008). In other words, people who have school accountability conception believe that assessment results show the school's success or even the country's success. Therefore, it is a way for schools or teachers to publicly show the quality of education they provide. Concerning school accountability, Wiliam (2010) states that students'

success was one of the critical factors determining the budget to be allocated to the schools in England in 1861. This was based on the Royal Commission report under the Duke of Newcastle. This understanding was a topic of debate, with some claiming that it was needed to make teachers and schools accountable by using an objective measurement basis. Some people, however, believed that it had an opposite and detrimental effect. On the other hand, Wickline (1971) discusses that this idea of accountability dates back to the times after the Second World War, when people were curious about the performance of schools.

There are different rationales behind school accountability conception. First, training provided by teachers and schools is labelled as “qualified” or “not qualified” based on the student's assessment results. To illustrate, it is claimed by Kane and Staiger (2001) that as of 2000, more importance started to be given to the test scores of schools, and rewards have been given to support them based on their success. Similarly, Figlio and Ladd (2015) indicate that there would be consequences of accountability, which might be positive or negative. In other words, bonuses or extra sources are spent on schools on the condition that they reach the expected level. Nevertheless, some actions and school restructuring become necessary when expectations are not met.

Another rationale for this conception is concerned with improving the quality of teaching and learning. As mentioned earlier, assessment not only indicates student performance but also gives teachers an idea about their teaching. This is why this view is related to creating qualified institutions and making necessary changes in teachers' perceptions. This might go hand in hand with the improvement conception. Therefore, assessments might help understand whether the teaching and planning phases have been practical and what students have learned. Additionally, Brown (2008) underlines that is a way for governments to ensure the quality of education and fund the schools, when necessary, as well as control teachers and see if the curriculum has been taught. Thus, high-stakes assessment, and in turn, school accountability create a unified basis and provide concrete results (Weeden, Winter & Broadfoot, 2002).

Nevertheless, this view of accountability may give rise to some problems. Firstly, with the idea of becoming successful in exams, teachers may focus on exam results and mainly use drill-based techniques. Likewise, announcing schools' results by labelling them as “good” or “bad” may deteriorate things in schools that cannot succeed in high-stakes exams. This competitive environment may cause the assessment to be perceived as irrelevant (Brown, 2008).

In summary, holding teachers and schools accountable for learning can have positive and negative results. For some people, providing incentives for schools or teachers due to the students' success might be positive and motivating, or it might be an excellent way to control teachers and schools. However, some may be negatively affected by the academic race, and it can make some teachers change their focus from teaching English to becoming successful in exams.

2.3.1.3 Student Accountability Conception

It is stated by Brown (2002) that “the premise of student accountability conception of assessment is that students are individually accountable for their learning through their performance on assessments” (p. 40). Therefore, it is characterised by assigning grades, checking students' performances with the help of a criterion, and placing them into different classes or schools, and students are accountable for their learning (Brown, 2004). Therefore, it is about making sure that students achieve the necessary qualities. This decision is helpful when placing them into higher educational institutions, or it is used as a part of the employment process by comparing them with their peers. It is concerned with whether students meet the objectives of the curriculum (Brown, 2008). Thus, teachers continuously get information about students with the help of assessment practices and interpret the data by considering other students' performances, objectives of the curriculum, or pupils' previous work. This strengthens the idea that students are individually accountable for their learning.

Kahn (2000) argues different reasons for assessment. These reasons for assessment are related to making students accountable for their learning through improving students' effort and motivation in the learning process. Also, the ones who are not responsible are penalised accordingly. In addition, with the changes in assessment practices, it is common for students to have a significant role in their assessment processes with self-assessment and peer assessment practices (Brown, 2008).

Some teachers may hold student accountability conception. However, some might be concerned about validity and reliability regarding high-stakes exams for vital decisions, such as student placement and employment. Moreover, Guthrie (2002) underlines student accountability conceptions causing negative emotions like anxiety and stress (as cited in Brown, 2008), which is affirmed by Berry (2008). S/he indicates that standards of high-stakes exams force students

to memorise answers, which harms learning in the long run and puts students and teachers under stress.

In conclusion, those with student accountability conception believe that assessment is a way to understand students' knowledge based on a criterion or compare them with other students' performances. They might also find it helpful to make students accountable for their own learning as a source of motivation even though there are opposite views, which may result in assessments being perceived as irrelevant.

2.3.1.4 Irrelevance Conception

The last conception is related to the idea that teachers' knowledge of students and the curriculum is enough to make deductions concerning the English levels of the students. Thus, "Assessment, usually understood as a formal, organized process of evaluating student performance, has no legitimate place within teaching and learning" as explained by Brown (2004, p. 305). Even though the ones holding this conception do not completely find assessment irrelevant, they believe that assessment should be intuitive with teachers' observations of students rather than formal exams, which have different rationales. For instance, assessment is seen as bad for teachers and students, which might be because assessment results are not true indicators of students' success or are ignored (Brown, 2004). Moreover, they may also indicate that assessment leads people to focus on other things, such as exams, rather than student learning, and it can even be detrimental for teachers in terms of having autonomy. This idea is supported by Berry (2008), who argues that high-stakes exams can have a negative backwash effect, diverting the focus from student learning to teaching for tests.

There are some other reasons for holding this conception. Firstly, it is highlighted by Brown (2008) that by focusing on students' emotions, which results from the student-centred education view, assessment is generally thought to be the same as testing, which prompts negative feelings for assessment. Furthermore, in his same study, Brown also includes Hawe's statements (2000) that students find assessment irrelevant because they do not believe that their papers are checked in a standardised way by all the teachers, making low achievers angry or demotivated (2008). This is supported by Struyven, Dochy, and Janssens (2005) who investigated students' beliefs, stating that luck plays an important role in end-point examinations with some other concerns leading to think that traditional assessment methods are unfair and inaccurate. Also, Stiggins

(1997) points out that not using assessment results to improve lacking points and teachers' insufficient knowledge and training about assessment might make them consider assessment irrelevant (as cited in Vardar, 2010). Finally, concerning the points related to previous assessment conceptions, it should be indicated that the pressure stemming from making students and schools accountable for learning can lead to assessment being considered irrelevant since the focus might turn to exams and exam results, rather than the learning process.

All in all, there are a range of reasons why teachers may think that assessment is irrelevant. Some might believe that assessment should be based on the intuitive judgement of teachers without stating that assessment is irrelevant. Besides, some teachers may have concerns about the psychology and well-being of students, or some may assert that teachers ignore assessment results and all those may lead to the irrelevance conception.

2.4 Summary

In summary, this chapter has defined assessment as one of the vital elements of teaching and learning. Different assessment practices performed in language classes have also been portrayed. Moreover, teachers' undeniable role in assessment and the importance of their conceptions and beliefs on assessment have been clarified. These clarifications have been supported by providing four assessment conceptions defined by Brown (2002, 2004, 2008) to make the study more understandable for the readers.

CHAPTER III

METHODOLOGY

In this chapter, the design of the study will be described. Then, research questions, context, and the participants will be presented. Finally, data collection procedure and instruments will be explained, followed by data analysis process and limitations.

3.1 Research Design

Conducting a research study entails knowledge and application of various research techniques. Dörnyei (2007) thoroughly discusses these different research methods. He defines quantitative research as mainly focused on numbers. This technique requires researchers to be fully prepared for categorisation and values before conducting the study. Also, quantitative studies deal with features in common in variables rather than highlighting individual cases, and statistics is an indispensable part of quantitative research, which paves the way for generalising the results by considering big numbers.

However, the nature of qualitative studies is much different. Qualitative research studies are flexible compared to quantitative studies. In other words, instead of pre-planning, categories and codes occur step by step during the research process. Moreover, the data is primarily verbal as the data collection tools are interviews and texts, or they are converted into texts to be analysed. In qualitative methods, participants' opinions and experiences are asked, making it inevitable to conduct the study with a small number of participants compared to quantitative studies. Lastly, it is worth noting that researchers have a significant position in qualitative studies as the data analysis is completed by researchers themselves based on their subjective perspectives instead of statistical programmes (Dörnyei, 2007).

A mixed-method research method following an explanatory sequential design was employed in this study. Creswell (2009) defines mixed-method technique as a combination of quantitative

and qualitative research elements. However, he underlines that it is not only collecting two different types of data but using and combining these two types simultaneously. Similarly, Teddlie and Tashakkori (2010) state that making inferences in mixed-method research studies includes thinking about how the findings respond to your research questions and using the quantitative and qualitative data results in an integrated to answer research questions. The reason for adopting a mixed-method design in this study is to answer the research questions comprehensively since the mixed-method design is an effective method to approach complex situations in a better way by obtaining numeric data from the quantitative research and combining them with the details gathered by the qualitative data, as indicated by Dörnyei (2007). Thus, researchers may favour using mixed-method research in cases in which using only qualitative or quantitative methods would not be sufficient (Creswell, 2009). In simpler terms, adapting qualitative methods to generalise the results to larger groups may not be possible. Also, by collecting the data with the help of the quantitative method, it is not possible to explore the participants' personal and detailed opinions. This makes using the mixed-method research design inevitable. In this study, explanatory sequential mixed method design was used by collecting and analysing quantitative data first, which is followed by the qualitative part since the qualitative part is connected to the quantitative part and is helpful for interpreting and explaining the findings gathered in the quantitative part (Ivankova, Creswell & Stick, 2006).

Fundamentally, the study aims to understand the assessment conceptions of English as a foreign language (EFL) instructors with the help of a scale and a focus group interview. Teachers' conceptions of assessment, which is the dependent variable in the study, is described based on four subscales: improvement, school accountability, student accountability, and irrelevance (Brown, 2002, 2004, 2008). Furthermore, while collecting data, the participants' demographic information was asked to answer some of the research questions. To this end, firstly, instructors' years of teaching experience, which is one of the independent variables, were asked to understand if assessment conceptions differ based on years of teaching experience. For this reason, the participants chose one of the following options given below:

- Less than two years
- Between 2-5 years
- Between 5-10 years
- More than ten years

Following this, the participants were asked to indicate their undergraduate departments to explain whether there is a relationship between the departments that the participants studied

and their assessment conceptions. With this aim, these six options were presented to the participants in the following way:

- ELT
- ELL
- Linguistics
- American Culture & Literature
- English Translation & Interpretation
- Other

After participants indicated their demographic information, they completed the TCoA-IIIA for the researcher to understand their assessment conceptions, a part of the quantitative method. Upon completing the scale, some participants, who were contacted based on their statements of willingness, participated in the focus group interview sessions, one of the most common qualitative research methods. The focus group interviews took place after the pilot study. While conducting different interview sessions, the researcher also took notes in her self-reflection journal to improve the process by reflecting on her steps and making the process transparent for the readers. These notes were included in the subsequent pages. Lastly, to analyse quantitative data gathered by the TCoA-IIIA, different tools, such as Analysis of Variance (ANOVA) or Kruskal-Wallis H test, were used based on results of the normality analysis. Furthermore, to analyse the qualitative data, the interviews were transcribed verbatim, coded with the help of MAXQDA, and placed under the four assessment conceptions defined by Brown (2002, 2004, 2008). The details regarding the study will be provided in the following pages of this chapter.

3.2 Research Questions

The study's main aim is to understand the assessment conceptions of EFL instructors. Moreover, their assessment conceptions will be associated with their years of experience and undergraduate departments to see whether there is a change in their assessment conceptions based on these two factors. Besides, the assessment practices used by the instructors will be investigated in detail. Based on these aims, this study will try to answer the following questions:

1. What are the assessment conceptions of EFL instructors in terms of four components defined by Brown (2002, 2004, 2008)?
 - 1.1. How do these conceptions vary according to instructors' experience years?

- 1.2. How do the assessment conceptions of the instructors differ based on their educational background?
2. What are the self-reported assessment practices that instructors apply in their institutions?
3. What is the relationship between EFL instructors' assessment conceptions and the self-reported assessment practices they apply?

3.3 Research Context

The study was conducted at the preparatory schools of six universities in Ankara/Türkiye. There were more schools than one in the data collection process to be able to gather more comprehensive information about the topic. The preparatory schools from which the participants attended in the data collection process were chosen based on convenience sampling. Some information concerning preparatory school programmes is necessary to comprehend the research context in more detail and is provided in the following paragraphs.

To begin with, a one-year preparatory school is offered to higher education students to teach basic English, which is supported by courses during their pre-service education. Considering the history of preparatory school programmes, different policies have been applied throughout the years (British Council & TEPAV, 2015). First, Boğaziçi University was the pioneer university in which a preparatory school was established in 1958, followed by Middle East Technical University (METU) in the 1960s. In 1996, however, the universities that provided English as a medium of instruction had to equip their students with a one-year preparatory school programme by offering an English for Academic Purposes (EAP) curriculum. Nevertheless, this was mandatory for the institutions using Turkish as a medium of instruction after 2001-2002. Later on, there was another regulation that made preparatory schools necessary only for the institutions that offer at least 30% of their courses in English, which led to some changes at institutions using Turkish as a medium since they wanted to retain their preparatory school programmes (British Council & TEPAV, 2015).

Moreover, by looking at the regulations published in the Official Gazette of the Republic of Türkiye (2016), it is possible to get a general idea of the preparatory school programmes of universities in Türkiye nowadays. First, an English proficiency exam is applied at the beginning of each school year for the students entering the university that year. This exam demonstrates

whether students are exempted from the preparatory school or not, along with the language proficiency level of the students. Preparatory school programmes aim to make students learn the basic rules of the target language, expand their vocabulary knowledge, improve their reading and listening skills, and make them eligible to express themselves verbally and in written forms. After completing the preparatory school year, students take another proficiency exam. If they fail the exam, they study in the preparatory school for one or two extra semesters. After that, they have to prove their language proficiency level by taking the exam at their universities or another English exam that their universities accept. In short, students are expected to pass the proficiency exam in two years at the latest.

Finally, it should be stated based on the same regulation in the Official Gazette that the English courses in the programme are taught by language instructors who are competent in the target language, which requires them to prove their language skills (Council of Higher Education, 2016).

3.4 Participants/Sampling

Researchers need to narrow down their number of cases by applying necessary sampling techniques. Hence, convenience sampling was used to select participants in the present study. Convenience sampling is defined as choosing participants because of their availability and easier accessibility for the researcher and participants' willingness to volunteer and participate. Thus, it is favoured by researchers more as it is an easy and inexpensive means of data collection (Dörnyei, 2007; Taherdoost, 2016). Mainly, accessibility accounts for the reason for using convenience sampling in this study since the researcher contacted schools that are easier to reach in Ankara.

One hundred and one participants from preparatory schools in various universities teaching in the 2022-2023 and 2023-2024 academic years in Ankara were selected based on convenience sampling to answer the TCoA-IIIA. Even though the initial aim was to achieve a larger number of instructors, the researcher encountered some limitations, such as the reluctance of some potential participants. Because of that, the total number of participants was limited to one hundred and one. Semi-structured focus group interview sessions were included as a follow-up stage to lessen the effect of this constraint and have a detailed understanding of instructors' opinions. With this aim, nine participants were chosen based on their willingness to participate in the focus group interview. In other words, the ones selected for this step had already left their

e-mail addresses to be contacted while gathering the data for the first part. Therefore, the researcher sent these instructors an e-mail for them to choose their available time slots (Appendix 4). After collecting this information, the researcher tried to include instructors with different experience years and different undergraduate departments in each session to have an effective discussion on the investigated subject in these three focus group interview sessions. Information concerning the participants of the study is provided in Table 2 and Table 3 below.

Table 2

Demographic Information of the Instructors (N=101)

Years of Experience	Number	Percentage
Less than two years	6	5.94 %
Between 2-5 years	15	14.85 %
Between 5-10 years	15	14.85 %
More than ten years	65	64.35 %
Departments	Number	Percentage
English Language Teaching	68	67.32 %
English Language & Literature	23	22.77 %
Linguistics	3	2.97 %
American Culture & Literature	3	2.97 %
English Translation & Interpretation	4	3.96 %
Other	0	0

As Table 2 demonstrates, the number of participants with less than two years of experience was small. Thus, the data analysis was conducted based on three categories rather than four: less than five years, between 5-10 years, and more than ten years. In addition, as the data gathered from the people who graduated from the Departments of Linguistics, American Culture & Literature, and English Translation & Interpretation was few, the data analysis included three categories again. These categories are ELT, ELL, and Other.

Furthermore, as included in the aforementioned paragraphs, some participants who wanted to participate in the second step were contacted for the focus group interviews, which took place in three sessions, after collecting the data in the quantitative part of the study. Detailed information about the participants taking part in the focus group interviews is presented in Table 3 below.

Table 3

Focus Group Interview Participants

Participants (Pseudonyms)	Session Number	Years of Experience	Undergraduate Department	Instructor at
Elif	Session 1 (I1)	More than ten years	English Language Teaching	University A
Esra	Session 1 (I1)	Between 5-10 years	English Language & Literature	University A
Leyla	Session 1 (I1)	Less than five years	English Language Teaching	University B
Ekin	Session 2 (I2)	Between 5-10 years	Translation & Interpretation	University A
Melda	Session 2 (I2)	More than ten years	English Language Teaching	University C
Ebru	Session 2 (I2)	More than ten years	English Language Teaching	University D
Hatice	Session 3 (I3)	Less than five years	English Language Teaching	University E
Buket	Session 3 (I3)	More than ten years	English Language Teaching	University C
Songül	Session 3 (I3)	Less than five years	English Language & Literature	University F

As shown in Table 3, there were four instructors with more than ten years of experience (Elif, Melda, Ebru, Buket) in the study. Moreover, while two instructors had experience years between 5-10 (Esra, Ekin), three instructors had less than five years of experience (Leyla, Hatice, Songül). Regarding the department, there were six instructors from the ELT Department (Elif, Leyla, Melda, Ebru, Hatice, and Buket). Furthermore, two instructors were graduates of the ELL Department (Esra, Songül). Lastly, one of the participants was a graduate of the Translation & Interpretation Department (Ekin).

3.5 Data Collection Procedure

First, three universities were chosen for data collection, and the necessary documents were provided to start the data collection process. Then, the researcher turned the scale into a Google Forms document by adding the necessary details, such as the demographics and information related to the study (Appendix 1 and Appendix 2). All the participants voluntarily participated in the study, gave their consent, and were asked to indicate if they were willing to participate in the following step. As the number of participants was insufficient initially, the researcher

visited all the schools in person to reach more participants and handed in the relevant documents. Nevertheless, the number of participants still needed to be increased. The researcher contacted the preparatory schools of three more universities (both public and private ones) in Ankara. After reaching one hundred and one participants, the second step, namely the focus group interview, was conducted by choosing nine participants from the ones eager to move on to the second stage after the pilot study, which included two instructors. Both the pilot study and the actual interview sessions were conducted online via Zoom. The interview sessions were recorded, transcribed verbatim, and analysed through coding with the help of MAXQDA. Lastly, the researcher took notes to ensure triangulation and improve her interview sessions and coding process as a self-reflection. Some entries from the self-reflection journal were also included to give readers explicit opinions of the researcher about the process (Appendix 5).

Table 4

Data Collection Process

Collecting Data with the TCoA-IIIA	May 2023- October 2023
Analysis of Quantitative Data	November- December 2023
Pilot Interview	November 2023
Collecting Data with Focus Group Interviews	December 2023
Researcher's Self-Reflection Journal	November 2023- March 2024
Transcription & Analysis of the Interviews	December 2023- March 2024

3.6 Data Collection Instruments

Various data collection instruments were utilised throughout the study to get answers from the participants and delve into the details of their conceptions.

3.6.1 Quantitative Data Collection Instruments

Firstly, a survey was used as a part of this study, which is an element of quantitative research design. Surveys are about researchers asking specific questions, collecting numbered data with the help of the participants, and analysing those numbers using statistics. Using a survey is an efficient way of collecting data as it helps researchers collect a considerable amount of data in a relatively shorter period without much cost. Therefore, a survey would be effective to discuss participants' background information and preferences (Creswell, 2012; McKay, 2006). Like many other research methods and practices, surveys have drawbacks along with their

advantages, one of which is superficial responses, as McKay (2006) expressed. Therefore, focus group interviews were conducted in the second phase of this study to minimise this handicap and have a more detailed framework of the participants' ideas, which will be explained below.

3.6.1.1 Teachers' Conceptions of Assessment Abridged Scale (TCoA-IIIA)

Brown explained the Teachers' Conceptions of Assessment Scale (CoA-III), which consists of 50 items, in his doctoral dissertation (2002). He proposed a four-facet model of teachers' assessment conceptions to comprehend the teachers' opinions concerning the nature of the assessment.

The shorter version of the scale, TCoA-IIIA, was used in the first part of this study because of a few reasons. First, the shorter version was developed by Brown later than the more extended version. Moreover, this version is even more efficient than the entire form, and similar information can be provided using the shorter one, as underlined by Brown (2006). TCoA-IIIA includes 27 items to be rated on a 6-point Likert scale from 1 (Strongly disagree) to 6 (Strongly agree) and four subscales: improvement, school accountability, student accountability and irrelevance.

Regarding this study, the researcher utilised this 27-item scale and did not feel a need to translate the scale into Turkish as all the participants are EFL instructors with an advanced level of English.

3.6.2 Qualitative Data Collection Instruments

Interviews, which can be encountered daily, such as on television or the radio, are among the most common ways of collecting data in qualitative research. Dörnyei (2007) indicates that conducting an interview means having a professional conversation with the interviewee with a purpose and a defined structure. Different types and designs that an interview can follow have been introduced in the literature.

In other words, the structure of interviews can range from highly structured to unstructured based on the questions asked (Dörnyei, 2007). Firstly, structured interviews leave no room for flexibility, and they are highly controlled with questions prepared beforehand. Hence, they have a similar logic as questionnaires. The researcher cannot provide any other information besides the questions prepared in advance. On the other hand, an unstructured interview follows a

direction with little interference from the researcher and is conducted with very little preparation. To make these two extreme ends meet, semi-structured interviews are the most commonly used interview type, as Brinkmann (2014) states. In semi-structured interviews, some questions are prepared beforehand. Yet, interviewees are asked to give details and explanations while being directed by the interviewer during the interview in this type, which allows researchers to take the best advantage of the interview process (Brinkmann, 2014; Dörnyei, 2007). A semi-structured interview was also integrated into the current study (Appendix 6 and Appendix 7).

3.6.2.1 Focus Group Interviews

Focus group interviews are applied by gathering 6-8 interviewees who are asked several questions and expected to offer their opinions, which is a relatively easy way to collect data in a short time by recording the answers of the group members (Dörnyei, 2007; Mckay, 2006). Gibbs (2012) points out that 4 to 12 people are generally suggested to conduct a focus group interview; however, there might be cases where this guideline is not followed. For instance, Toner (2009), who preferred including two people in her focus group study, underlined in her paper “*Small is not too small*” that small focus groups follow the same structure as large groups by providing thick and detailed data. Therefore, it can be concluded that there is no agreement on a specific number when it comes to focus group interviews. While groups involving more people can be complex to monitor and moderate by the interviewer, too small groups risk not providing a comprehensive discussion (Swartling, 2007).

Conducting focus group interviews can be beneficial, as Creswell (2007) expresses. He states that focus group interviews are most likely to be the best information source if similar participants cooperate; however, if there is a time limit, participants might feel more hesitant to give information. Moreover, researchers’ role in this type of interview is to be the leader and the facilitator of the discussion by ensuring that everybody participates equally (Dörnyei, 2007).

In light of the information above, a focus group interview was conducted as a follow-up stage to understand the participants' opinions comprehensively. Interviewees were asked 15 semi-structured interview questions prepared by the researcher (Appendix 7). After the pilot study, which took approximately 60 minutes and had only two participants, the researcher decided to conduct the focus group interview in three different sessions by including three participants in each one. The reason for this choice was to help participants express their opinions more in

detail without facing any concerns that might occur due to including 7-9 people in one session simultaneously. All the interview sessions were conducted via Zoom and video-recorded, which paved the way for the transcription and analysis.

Table 5

Duration of the Interview Sessions

Focus Group Interviews	Duration
Session 1	72 minutes
Session 2	59 minutes
Session 3	65 minutes

After the first interview session, the researcher thought that having a session of about one hour with three participants was effective and the session provided comprehensive data, so the other interview sessions were conducted in the same way, as stated in the researcher's self-reflection journal.

I think including three people in one focus group interview was effective. Even though I, at first, felt like people were affected by each other's ideas, they could elaborate on what they thought, so it was okay. As it took 72 minutes, including more people would have caused boredom, or "I agree" statements would have been more, maybe. That is why I think the first session was highly effective. (Journal entry, December 12, 2023)

Overall, three semi-structured focus group interview sessions were conducted, and they provided in-depth data concerning the participants' opinions on assessment and their assessment practices.

3.6.2.1.1 Pilot Study

The importance of conducting pilot studies is underlined by Majid, Othman, Mohamad, Lim, and Yusof (2017), who underline the difficulty of conducting interviews for inexperienced researchers. With the help of the pilot studies, researchers can prepare for the actual interview by adjusting the required points, leading to a quality interview process.

To ensure a smooth and quality interview process in this paper, first and foremost, the researcher asked for expert opinions on the interview questions. For that reason, three experts working at the ELT Programme of Gazi University were requested to share their views and suggestions on the semi-structured interview questions prepared by the researcher. After getting their ideas on the questions, necessary changes were applied. In other words, the wording of some questions was changed to make them more understandable for the participants.

After that, a pilot study was conducted with two instructors before the focus group interview sessions to see whether the questions needed further revision or different wording. The participants of the pilot study were chosen based on convenience sampling. They had different educational backgrounds and years of experience. The first participant, who had experience of five years, graduated from the ELT Department. In contrast, the other participant had seven years of experience and graduated from the Department of Translation & Interpretation. Before the pilot study, they were informed about the aim of the interview and consent forms were sent to them. Also, they were kindly asked to share their feedback on the interview process and the interview questions after completing the pilot study. Following the pilot study, the researcher made some decisions, as written in her self-reflection journal, and some changes were applied.

I am happy that my friends could express their opinions without hesitation; however, I have other concerns now. This interview took almost one hour with two participants, so how am I going to include 7-9 participants in the same interview, and how long will it take? Including a small number of participants may make the discussion more effective. Also, I have realised that some questions needed more explanations and examples. Lastly, I think I can add some more questions about Brown's assessment conceptions because I have realised that I have not included specific questions concerning two conceptions defined by him. (Journal entry, November 30, 2023)

As stated above, after witnessing that the pilot interview lasted for an hour with two participants, the way the actual interview would take place was changed so that everyone could express their thoughts and experiences more in detail without time limitations. Secondly, some questions were decided to be asked by providing examples for the interviewees. For instance, the question “*What assessment practices do you use to see your learners' progress?*” was asked by giving examples to create a clear framework in the participants' minds. Last but not least, a few questions were added to refer to the assessment conceptions defined by Brown, whose scale was applied in the first step of the study. This led to making better inferences concerning the topic and research questions. For example, the question “*Is there a relationship between the success of teachers or institutions and students' assessment results? How?*” was included in the actual interview, referring to the school accountability conception defined by Brown (2008).

3.7 Researcher's Self-Reflection Journal

Although it did not serve as a data collection tool, the researcher kept a self-reflection journal in the data collection and analysis phases of the qualitative data so that the researcher could reflect on her own opinions and decisions. Similarly, Ortlipp (2008) points out that the reflective approach is a common technique in qualitative research studies and researchers have a space to elaborate on their perspectives, experiences, and assumptions by including more transparency

in the research study by using reflective journals. In the same article, the writer aims to help novice researchers integrate reflective journals into their research papers by expressing the scarcity of those research papers in the field. Despite being scarce, it is an effective method by which researchers can examine their own beliefs and foster self-awareness. For instance, Ortlipp (2008) explains how s/he changed some of his/her preferences during the research process because of having a chance to reflect on his/her choices with the help of the reflective journal.

In this vein, Watt (2007) highlights some benefits of using reflective journals for novice researchers by stating that they are vital in becoming better researchers. Firstly, it helps researchers to comprehend their own ideas. In the same article, she claims that she would have been lost if she had not written down her ideas. In addition, reflective journals open a window for readers to see and understand the opinions of researchers, which might be effective in understanding the study better. The researcher in this study also included some entries from her self-reflection journal to the study to make the process more transparent for the readers and add credibility to the research study.

3.8 Data Analysis

3.8.1 Analysis of Quantitative Data

Firstly, the responses to the TCoA-IIIA, which is a 6-point Likert scale, were collected to analyse the data. As some of the participants were given sheets with the scale and demographic information by hand and their responses were not gathered via Google Forms, they were added to the Excel sheet by the researcher for the following steps concerning the analysis of the quantitative data. The TCoA-IIIA was analysed for validity and reliability purposes through Confirmatory Factor Analysis (CFA) and Cronbach's Alpha. Furthermore, the frequency distributions of the responses were written based on the participants' experience years and undergraduate departments. Based on that, some decisions were made. In other words, the demographic information was presented in options for participants to choose from. Firstly, the experience years of the participants were divided into four categories, and the participants chose the one that explained their situation in the best way. The options are given below.

- Less than two years
- Between 2-5 years

- Between 5-10 years
- More than ten years

Nevertheless, because the number of participants was a limitation, two of the categories were merged to reach more meaningful results. Thus, the analysis was done based on the three following categories.

- Less than five years
- Between 5-10 years
- More than ten years

Moreover, the participants were asked to choose the department that they had graduated from based on the six options shown below.

- ELT
- ELL
- Linguistics
- American Culture and Literature
- English Translation & Interpretation
- Other

In the same vein, while checking the respondents' answers, it was decided that the data would be analysed in three categories due to the limitation in the number of participants in some options shown above. The three categories included in the data analysis part are presented below.

- ELT
- ELL
- Other

The aforementioned categories, along with the frequency distributions of respondents are demonstrated in Tables 8 and Table 9 below.

Table 8

Frequency Distributions Categorised by Experience

How long have you been teaching?	Frequency	Per cent
Less than five years	21	20.8
Between 5-10 years	15	14.9
More than ten years	65	64.4
Total	101	100.0

Table 9

Frequency Distributions Categorised by Department

Which department did you graduate from?	Frequency	Per cent
English Language Teaching	68	67.3
English Language & Literature	23	22.8
Other	10	9.9
Total	101	100.0

As Table 8 demonstrates, instructors with more than ten years of experience constitute the bigger part of the study (64.4%), followed by participants with less than five years of experience (20.8%). Lastly, participants with 5-10 years of experience constitute the smallest portion of the participants (14.9%). On the other hand, Table 9 shows that the participants are mainly graduates of the ELT Department (67.3%). This is succeeded by the ELL graduates (22.8%). Some other instructors graduated from other departments, such as Translation & Interpretation, American Language & Literature, and Linguistics. Since the number of these participants was small, they will be evaluated under the Other category (9.9%).

Following this categorisation, descriptive statistics were conducted and an overall idea was shaped. After that, the data was analysed more specifically. In other words, the assessment conceptions of the instructors were studied based on their years of experience first. The distribution of the responses was checked via the Shapiro-Wilk Test to see normality and ANOVA was applied. Then, the analysis was conducted to see the differences concerning assessment conceptions of instructors based on the department they graduated from. The Kruskal-Wallis H test, which is a nonparametric test, was applied in this step as the normality assumption was not met, as demonstrated by Shapiro-Wilk Test results. All the analyses in this step were run with the help of statistical programmes, such as SPSS and SPSS AMOS.

3.8.1.1 Validity and Reliability

Quantitative data collection can be gathered via surveys. In that case, there are certain points to be considered regarding the data collection instrument, such as validity and reliability. To begin with, validity is about making meaningful deductions from scores on the data collection instrument. Thus, ensuring validity is significant since it gives an idea about the suitability of the data collection instrument in survey research. Three different types of validity are explained by Creswell (2009) in the following way: Firstly, content validity is concerned with whether the content is effectively measured by the items on a scale. Secondly, concurrent validity is about the correlation of results with other results. Lastly, Creswell (2009) explains construct validity and expresses that it is about whether theoretical concepts are measured by the items and whether scores are beneficial and have a functional purpose when they are applied in practice.

On the other hand, reliability is checked to understand whether consistent results can be achieved when a survey is conducted at different times or whether consistency is achieved in test administration and scoring phases (Creswell, 2009).

In the present study, the TCoA-IIIA was used to collect data, and validity and reliability analyses were conducted. Cronbach's Alpha was used to check reliability, and CFA was applied to check validity. The results of the validity and reliability analyses of the TCoA-IIIA are presented below.

3.8.1.1.1 Validity Analysis of the Scale

The results of the CFA indicate that the model is structurally valid, and the data adequately fits the model that consists of 27 items and four sub-dimensions (fit indices: $\chi^2/df= 2.30 < 5$). Nevertheless, other fit indices RMSEA, SRMR, and CFI values are 0.114, 0.136, and 0.680, respectively. Thus, they are beyond the proposed thresholds of $0.05 < RMSEA < 0.05$, $SRMR < 0.10$ and $CFI < 0.90$. This is mainly because the sample size is not large enough. Therefore, this limitation should be considered regarding the outcome of CFA.

As demonstrated in Table 7, the relationships between all items and the subscales are statistically significant (p-values < 0.05).

Table 6

CFA Results

Standardised Measurement	Satorra-Bentler Coef.	Std. Err.	z	p> z	[95% Conf. Interval]	
Item 2	.4151991	.0584228	7.11	0.000	.3006925	.5297057
Student Acc.	3.807026	.2250177	16.92	0.000	3.365999	4.248052
Item 11	.4798612	.0577812	8.30	0.000	.3666121	.5931103
Student Acc.	3.357235	.1990962	16.86	0.000	2.967014	3.747457
Item 20	.7740417	.049092	15.77	0.000	.6778232	.8702603
Student Acc.	4.349307	.2422	17.96	0.000	3.874604	4.824011
Item 1	.672451	.0376479	17.86	0.000	.5986625	.7462395
School Acc.	3.927898	.231849	16.94	0.000	3.473482	4.382313
Item 10	.7446201	.0421136	17.68	0.000	.6620791	.8271612
School Acc.	3.391323	.1669496	20.31	0.000	3.064108	3.718538
Item 19	.86439	.0239457	36.10	0.000	.8174573	.9113227
School Acc.	3.027196	.164187	18.44	0.000	2.705395	3.348997
Item 7	.4577643	.0945253	4.84	0.000	.2724982	.6430304
Irrelevance	2.391984	.1295301	18.47	0.000	2.13811	2.645858
Item 8	.5551368	.0641355	8.66	0.000	.4294334	.6808401
Irrelevance	2.715144	.1614849	16.81	0.000	2.39864	3.031649
Item 9	.2698692	.0870639	3.10	0.002	.0992271	.4405114
Irrelevance	3.400385	.1848838	18.39	0.000	3.03802	3.762751
Item 16	.6210119	.0605962	10.25	0.000	.5022455	.7397783
Irrelevance	2.12008	.1133838	18.70	0.000	1.897852	2.342308
Item 17	.7364711	.0445871	16.52	0.000	.6490821	.8238602
Irrelevance	1.935016	.1276221	15.16	0.000	1.684881	2.185151
Item 18	.2953318	.0837434	3.53	0.000	.1311978	.4594658
Irrelevance	4.069846	.2314912	17.58	0.000	3.616131	4.52356
Item 25	.5592866	.0572755	9.76	0.000	.4470287	.6715446
Irrelevance	2.203481	.1377238	16.00	0.000	1.933547	2.473414
Item 26	.3367285	.0884267	3.81	0.000	.1634153	.5100417
Irrelevance	2.101711	.1117074	18.81	0.000	1.882769	2.320654
Item 27	.6110161	.0648408	9.42	0.000	.4839304	.7381018
Irrelevance	2.19229	.144623	15.16	0.000	1.908834	2.475746

Table 6

CFA Results (Continued)

	Standardised Measurement	Satorra-Bentler Coef.	Std. Err.	z	p> z	[95% Conf. Interval]	
Item 3		.7379038	.0272967	27.03	0.000	.6844032	.7914044
Improvement		4.598957	.2914797	15.78	0.000	4.027668	5.170247
Item 4		.6919229	.0352825	19.61	0.000	.6227705	.7610752
Improvement		4.443907	.2580246	17.22	0.000	3.938188	4.949621
Item 5		.6604882	.0401267	16.46	0.000	.5818413	.7391352
Improvement		4.756567	.3108193	15.30	0.000	4.147373	5.365762
Item 6		.6029707	.0546259	11.04	0.000	.4959059	.7100355
Improvement		4.455828	.2685926	16.59	0.000	3.929396	4.98226
Item 12		.8400856	.0167038	50.29	0.000	.8073468	.8728244
Improvement		3.853133	.2255451	17.08	0.000	3.411073	4.295194
Item 13		.8431082	.0205256	41.08	0.000	.8028786	.8833377
Improvement		4.236934	.2236632	18.94	0.000	3.798562	4.675306
Item 14		.7394257	.0354458	20.86	0.000	.6699532	.8088981
Improvement		4.079674	.2035561	20.04	0.000	3.680712	4.478637
Item 15		.5352082	.0409573	13.07	0.000	.4549334	.6154829
Improvement		3.515282	.1715374	20.49	0.000	3.179075	3.851489
Item 21		.5375201	.0492405	10.92	0.000	.4410105	.6340296
Improvement		2.931183	.152012	19.28	0.000	2.633245	3.229121
Item 22		.7242483	.0334203	21.67	0.000	.6587456	.789751
Improvement		4.368726	.2073424	21.07	0.000	3.962342	4.77511
Item 23		.4059325	.0508975	7.98	0.000	.3061752	.5056898
Improvement		2.372996	.1367384	17.35	0.000	2.104994	2.640999
Item 24		.6613562	.0442116	14.96	0.000	.574703	.7480094
Improvement		3.790939	.2188913	17.32	0.000	3.36192	4.219958

3.8.1.1.2 Reliability Analysis of the Scale

To predict the reliability of a psychometric test, Cronbach's Alpha is used, which can either be calculated for each scale item or as the average of all the items in the scale (Kılıç, 2016). Therefore, Cronbach's Alpha is used to see the relationship between a set of items as a group, and it helps decide if a score is reliable. Thus, it was implemented to understand the internal consistency of the scale in this study. While deciding the reliability score, Cronbach's Alpha of 0.70 and above is considered good, while 0.80 and above is better. Finally, Cronbach's Alpha of 0.90 and above means the best.

In the present study, the overall reliability coefficient of the TCoA-IIIA is 0.863, which means that the scale has good internal consistency. Moreover, the subscales in the scale are also satisfactory. The Cronbach's Alpha value of the subscale student accountability is 0.603, which is not low but can be considered questionable.

Table 7

Reliability Statistics Results

	Cronbach's Alpha	N of Items
TCoA-IIIA	0.863	27
School Accountability	0.790	3
Student Accountability	0.603	3
Improvement	0.900	12
Irrelevance	0.749	9

3.8.2 Analysis of Qualitative Data

Conducting a qualitative research study and analysing the data differ from quantitative data collection and analysis practices. While quantitative data mainly deals with numbers to collect a vast amount of data in a shorter period, qualitative data focuses on text or image data (Creswell, 2009) to comprehend the details of a phenomenon. Dörnyei (2007) describes qualitative analysis as one based on language. He supports this idea by stating that the audio recordings are also transcribed into textual form for analysis. Therefore, the analysis in the qualitative research studies is mainly language-based. Thus, that the aim of qualitative data is generally to examine a topic from a deeper perspective with the help of the opinions and experiences of the participants. In this regard, the qualitative data was analysed through qualitative content analysis in this study after the data collection phase.

When analysing the data gathered through qualitative techniques, researchers interpret their data, which might be based on their beliefs, science, or both (Creswell, 2009). A careful analysis is required by following some specific steps in qualitative analysis, which Creswell defines as peeling back the layers of an onion (2009). Although some variations might be defined in the literature, Creswell (2007) states specific steps that researchers follow in qualitative content analysis.

The first step of qualitative content analysis is the preparation and organisation of the data, as also mentioned by Dörnyei (2007). In this step, the data acquired by qualitative methods are

transcribed and organised for the following steps. Transcription and data organisation allow for data reduction through creating themes and codes. Finally, after understanding and demonstrating logical links in the data by codes and themes, the results are discussed and presented in tables and figures (Creswell, 2007). A more comprehensive table about the steps of qualitative data analysis is provided below.

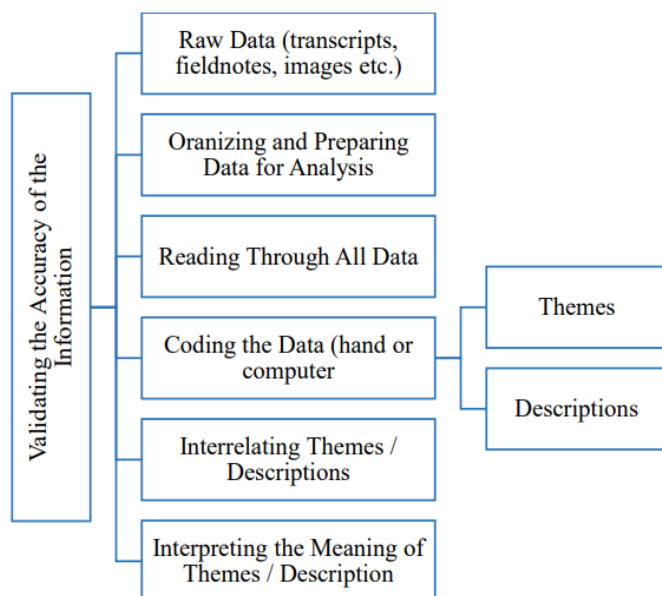


Figure 2. Qualitative data analysis. Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. The United States of America: SAGE.

Unlike quantitative analysis, qualitative content analysis does not follow a linear process (Dörnyei, 2007). In simpler terms, researchers always go backwards and forwards in qualitative analysis while collecting, analysing, and interpreting the data. This can occur, for instance, when researchers witness emerging information and code the data accordingly based on these new findings. Creswell (2007) also demonstrates this iterative process.

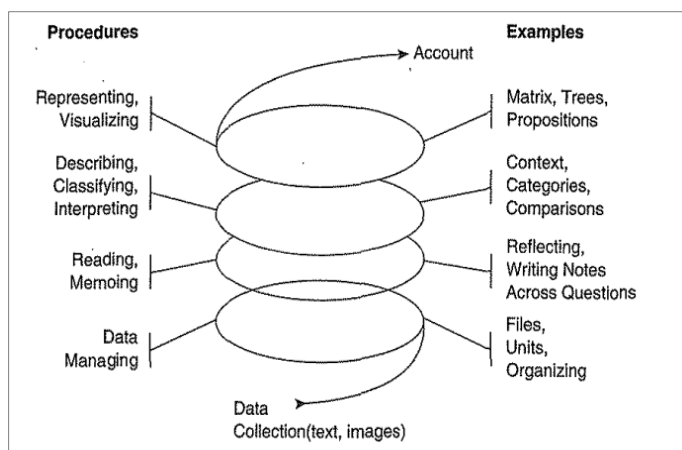


Figure 3. Qualitative data analysis spiral. Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches*. The United States of America: SAGE.

Moreover, some significant points regarding the coding procedure in qualitative analysis should be considered, as Creswell (2007) argues. Firstly, whether to use pre-existing codes is controversial in the research literature. While Creswell (2007) claims that using pre-existing codes is generally encountered in health sciences, he suggests that it is advantageous to adapt and include emerging codes during the data analysis even when researchers use pre-existing codes. In this study, the researcher detected recurring statements and clarified the codes that occurred during the data analysis process and interpreted these codes under the four pre-defined themes explained by Brown (See Chapter II).

Secondly, some researchers are in favour of counting the number of codes during the analysis. In contrast, others may not prefer this practice since this is more of a quantitative way to analyse the data, as discussed by Creswell (2007). In the present study, the codes were not counted or presented in tables while the data analysis was being reported.

Lastly, another controversial issue in qualitative analysis concerns the names given to the codes during the coding process. Researchers may follow different paths while coding their data. For instance, some researchers might prefer to label the codes with the exact words of the participants, which is called in-vivo coding; however, some might prefer labelling their codes with the words that describe their cases in the best way. As Creswell (2007) suggests, the researcher in this study tried to include codes that the researcher expected before the study as well as the information that was interesting or unfamiliar during the coding process. Also, these codes were named by the researcher rather than in-vivo coding. The codes and the results of the analysis will be presented in Chapter IV.

Based on the information above, the data in this study was analysed by following the steps defined by Creswell (2007). First, the researcher checked the video recordings of the interview sessions and transcribed the data verbatim. The transcription process and the observations of the researcher are exemplified in the following way:

I have been transcribing the data that I gathered from three different groups of participants. Right now, I feel like they were focused on different points while answering my questions. I mean, I could catch some common frameworks, I think. However, I also think that it may be more difficult for me to come up with recurring statements because the instructors had different profiles, and there were participants who had been working in the testing office of their schools, so they focused on the testing and exam parts more than the first participant group. The recurring ideas that I could detect while transcribing are given below:

1. *Participants believe that assessment is crucial.*
2. *The schools use a variety of techniques to assess their students; however, some revision is needed, for example, to contextualise the assessment practices and to make them more authentic.*
3. *Teachers agree that their undergraduate courses were not enough to shape their understanding of assessment and that they had to learn things while performing their jobs. Moreover, some of their ideas on assessment have changed over time with more practice. (...)*
(Journal entry, January 21, 2024)

After that, the researcher read and reread the transcriptions to get familiar with the content and not miss any significant points. After completing these steps, the data was interpreted into codes under this study's four pre-determined themes: improvement, irrelevance, student accountability, and school accountability (Appendix 9). Codes are defined by Saldaña (2013) as a part of qualitative data, i.e. data based on visuals or language, like a word or a phrase, giving the essence or summary of the meaning. The researcher kept this in mind and coded and interpreted the data accordingly. Creswell (2009) indicates that the data interpretation could be the researcher's personal interpretation based on his/her experiences, or the meaning might be obtained based on the comparison with the existing literature. The researcher in this study interpreted the data and discussed it by comparing it with the literature findings (See Chapter IV). While analysing the qualitative data, MAXQDA, which is a software programme that aims to help researchers in coding their qualitative data, was used since it is time-saving. Moreover, it allowed the researcher to code their data more easily with the help of its effective interface.

3.8.2.1 Credibility and Trustworthiness

Saldaña (2011) explains the terms credibility and trustworthiness in the collection, analysis, and presentation of qualitative data. In other words, researchers are required to provide rationales for their preferences concerning the number of participants, methodology, data analysis, and conclusion parts to convince the readers to trust the research study (Saldaña, 2011), and there

are certain ways to achieve credibility in qualitative studies. McKay (2006) indicates that researchers should transcribe and analyse the data attentively without bias to ensure credibility. Lincoln and Guba (1985) also list some techniques to be applied to ensure credibility, as presented below.

1. Engagement in the related field,
2. Constant observation,
3. Utilisation of many different research methods,
4. Having discussions with peers about the study design,
5. Member checking (as cited in McKay, 2006).

Similarly, Creswell (2007) suggests eight ways to maintain validation (as he calls it), some of which are parallel to Lincoln and Guba's steps. Moreover, he suggests using two of these techniques in qualitative research studies.

In this research study, the researcher applied some of the recommended stages. First of all, as Creswell (2007) points out, the external check of the research study process is about peer review, similar to interrater reliability included in quantitative studies, and peer review was utilised in this study. To this end, the data and the codes were checked by an expert working at Gazi University. Secondly, triangulation was attempted by using different research methods to comprehend the investigated phenomenon deeply. Moreover, prolonged engagement, which is about being familiar with the investigated community, was also valid for this research study since the researcher is a colleague from the same field as the participants. Finally, member checking was conducted by asking one of the participants in the focus group interviews to check the codes explained by the researcher to avoid any possible misunderstandings. In addition, trustworthiness or credibility is acquired through informing the readers about the research process (Saldaña, 2011). Therefore, the researcher tried to provide an in-depth description of the context and the participants by supporting the findings and reasons for her choices with her self-reflection journal, which Saldaña recommends keeping throughout the coding process of the qualitative data. He calls these notes "analytic memos" (2013, p. 41). While writing analytic memos, researchers can think critically and reflect on their own thoughts and choices, which was achieved in this study.

3.9 Limitations of the Study

This study has some limitations that need to be addressed. The first limitation is the number of the participants. To put it differently, even though the researcher aimed to reach more participants, some factors, such as the reluctance of some potential participants, led to the number of participants for the questionnaire being limited to one hundred and one instructors. This limited number should be taken into account while interpreting and generalising the results of the study. Including different preparatory schools and more instructors could have made this study more comprehensive and easier to generalise the results to the whole population. Secondly, the findings concerning participants' assessment practices are based on their own reports in the focus group interview sessions rather than real-life classroom observations. Thus, observations regarding participants' assessment practices can be added to the future studies as another data collection tool.

CHAPTER IV

FINDINGS AND DISCUSSION

In the previous chapters, the introduction, literature review, and methodology of the study were comprehensively explained. This chapter will provide the findings gathered through quantitative and qualitative data collection instruments and these findings will be discussed in the light of the literature and the research questions of the present study.

4.1 RQ1: What are the assessment conceptions of EFL instructors in terms of four components defined by Brown (2002, 2004, 2008)?

Both quantitative and qualitative parts of the study helped understand the answer of the first research question. First, descriptive statistics summarised quantitative data. Mean scores and standard deviation were calculated and the information concerning the four subscales of TCoA-III A is represented below.

Table 10

Descriptive Statistics Results

Subscales	Minimum	Maximum	Mean	Std. Deviation
Student Acc. Score	6	18	12.64	2.496
School Acc. Score	4	18	11.55	2.837
Irrelevance Score	14	43	30.13	6.876
Improvement Score	22	70	50.73	9.106

As demonstrated in Table 10, improvement conception (M=50.73) has the highest mean score and agreement level, followed by irrelevance conception (M=30.13). Even though student accountability (M=12.64) is the third one agreed upon, the mean score of school accountability conception (M=11.55) is not very different from the student accountability conception. Since there was a need to comprehensively understand instructors' opinions on assessment and

assessment practices as well as the reasons for their assessment conceptions, focus group interviews were conducted with instructors working at preparatory schools of six different universities in Ankara/Türkiye. While choosing the participants, the researcher tried to include instructors whose educational backgrounds and experience years vary as much as possible to have a better discussion in each session. Focus group interview sessions yielded rationales behind the findings gathered through descriptive statistics, and the findings are presented and discussed below.

4.1.1 Improvement Conception

As descriptive statistics results suggested above, improvement conception ($M=50.73$) has the highest mean score and agreement level in this study. This finding resembles Brown's findings (2002). He found that one of the most agreed conceptions by the teachers in his context was the improvement conception. Moreover, the study conducted by Brown and Chen (2015) in the Chinese context concerning the assessment conceptions of teachers also revealed that improvement conception had the highest value, as Ma and Bui (2021) also argued. In the same vein, investigating Chinese teachers' conceptions of assessment in their own study, Ma and Bui (2021) concluded that the participants mostly agreed with the idea that assessment was used to improve learning. Finally, Azis (2012) analysed many assessment conceptions studies in his meta-analysis study. Remarkably, s/he concluded that teachers mainly favoured improvement conception regardless of their cultural differences. The present study reached a similar conclusion about the main conception of EFL instructors held in Ankara/Türkiye in the preparatory school context at the tertiary level. Improvement conception's having the highest mean value may indicate that EFL instructors working at preparatory schools are aware of the significance of assessment and believe that assessment helps improve the language teaching and learning process, which is parallel to the findings of Pastore and Mincu (2024) in the Italian context. Some statements of the participants addressing improvement conception and the related codes are presented below.

4.1.1.1 Essentiality

First of all, the participants in this study believe that assessment is a necessary tool for teaching and learning. Initially, Elif thinks that assessment is a map to guide her since she gets help from

assessment results to revise things or build up her future practices. However, she does not label assessment as the most significant element in language teaching.

[...] Assessment, at some points, is necessary to see where we are to go: forward or maybe to go back to fix something or to add something more to their learning. But it is not the most important thing. (Elif, I1)

Quite similarly, Ebru describes what assessment is from her own perspective and expresses the importance of it in undertaking the necessary actions.

For me, assessment is useful for getting feedback from students, especially because you know what is conveyed and what is not conveyed to the students after you assess something. [...] (Ebru, I2)

[...] As a teacher, I must know where I stand and where my students stand. So, if I want to focus on the lesson, I have to do assessments and apply some assessment strategies in my classroom, so I need to keep up with the lesson and the students. (Ebru, I2)

While explaining her thoughts, Melda explicitly expresses that assessment strategies have one primary purpose: improving the students' learning. Besides, she includes other related elements, such as testing and certification, which will be mentioned under the student accountability conception. In the first part of the excerpt below, Melda states that she is a member of an organization named METASIG and explains its meaning. After that, she indicates her agreement with Ebru's aforementioned statements.

[...] Our name is META; measurement, assessment, testing, and evaluation because assessment includes all of them. We do have numbers in it. We do have our evaluations in it, and as my colleague mentioned, we use all those things for just, I think, one purpose, which is to improve and contribute to student learning. But at the same time, we must do it for certification and decision-making purposes, so it is a concept that has many, many, many layers and aspects. (Melda, I2)

Instructors need to include assessment practices in the teaching and learning process to understand whether learning has happened, which might lead to reconsidering their ongoing teaching practices. Similarly, the other participants agree upon the necessity of assessment.

It is really important, and it is a need. Both students and teachers need assessment because students also need to see what they can or cannot do. Therefore, it also shows the path to teachers and students, which is very important. (Esra, I1)

Moreover, Ekin states that she has been working in her institution's testing office and describes the importance of assessment from both a teacher's and a testing office member's perspective.

[Assessment is] like aligning with objectives. We can say that it is an ongoing process. It has a lot of steps, and each one is essential, from choosing what to ask to disseminate the results. It is a long journey, and each step is crucial and valuable. (Ekin, I2)

Yeah, sure, it is [important]. For me as a tester, [it is important] to see if the curriculum is working, to understand the needs of the students, to see what else we can do. As a teacher, [assessment is essential] to keep track of my students' progress, to see if I can help them in any other way, and to see if there is something that I could not cover in my class. (Ekin, I2)

Finally, Leyla highlights the importance of assessment with the following words. She shares Elif's opinion, stating that assessment helps her decide her future practices. In other words, based on students' performance, Leyla considers providing them with extra studies.

[...] I think the most important part of [assessment] is that it helps us move forward or go backward and revise things, which is very, very important, at least to me. (Leyla, I1)

In short, assessment is defined as a significant component by the participants of this study, which is evaluated under the improvement conception.

4.1.1.2 Assistance for Teachers

In the first section, instructors describe what assessment is and highlight its importance. In this section, the main idea is that assessment helps teachers to get information about students and take necessary steps accordingly. However, it should be noted that the importance of getting information about students can stem from many different points. Firstly, Leyla indicates getting an idea about her learners' psychological needs based on the assessment results.

[...] Apart from the classroom setting and individual interactions with students, it is the only way you can check how a student is doing, not only in terms of academic success but also had the opportunity to see that a student is having a hard time emotionally, for example, just by looking at the consistency of their test behaviour, etc. [...] (Leyla, I1)

In a similar vein, Hatice believes that assessment not only provides information about her students' knowledge and performance but also helps her understand their different needs.

[...] Assessment not only assesses learners' mechanic knowledge of English but also it should assess pedagogically; maybe learners' needs and how they react to learning a new language in a classroom atmosphere because there are many individual differences when we observe a classroom... So, it not only assesses learners' exam results but also assesses their psychological needs. (Hatice, I3)

As demonstrated above, different from the results of Zaimoğlu (2013), Leyla and Hatice mention that assessment helps them to get ideas about students' psychological needs, along with their academic needs, which is an interesting point to be underlined. Moreover, Elif believes that assessment practices assist her in evaluating herself as a teacher, which is like a reflection of her teaching.

[...] I guess there are two ends to the assessment. First, we need to see whether we are successful, assess our teaching, and see the learner side, such as how much they learned and the weak and strong parts of the topic, let us say. (Elif, I1)

Similar to Leyla and Elif, Ebru also gets help from assessment techniques to decide on her further actions as a teacher.

For me, assessment is useful for getting feedback from students, especially because you know what is conveyed and what is not conveyed to the students after you assess something. You can decide on the ongoing parts of the progress depending on your assessment, so, it always happens during the classroom in each minute, maybe. (Ebru, I2)

Last but not least, Songül mentions how she acquires assistance from assessment results. She focuses on the reasons for her students getting good or bad grades, for example, after an exam. Spending some time evaluating students' exam results and considering the underlying reasons

for their failure or success might pave the way for her to decide whether she needs to make changes in her teaching style or whether she can do other things to help her students.

[...] Let us say that the average is 10. Then, I categorise students into five or four groups like a band system: 9 and 10 will be a band, 12 and 13 will be a band, and so on. So, I see how many students are in which band, and I would like to focus on the reasons. (Songül, 13)

In short, the statements that instructors point above concerning the improvement of teaching and learning align with Brown's improvement conception, which is based on the idea that assessment helps enhance teachers' practices and, in turn, student learning (Brown, 2002). In a similar vein, Lutovac and Flores (2022) found that PSTs mostly think that assessment should inform instruction and teaching together with the learning that would take place in the future. Therefore, it can be said that both PSTs and in-service EFL instructors share a similar point of view. Also, EFL instructors' other opinions concerning how they use assessment results are provided below.

4.1.1.3 Guidance

As included in the abovementioned excerpts, instructors get help from assessments to improve their current assessment practices. Because of that, assessment results are not "filed and ignored" as mentioned in one of the items of TCoA-IIIA, related to the irrelevance conception of assessment. Firstly, Buket, one of the most experienced teachers in the interview, exemplifies making use of assessment results in the following way.

I have an example of that. Once, most of the students in my class were not able to get the difference between "mustn't" and "don't have to." Then, I realised that I must have done something wrong while I was teaching that part, and I went over that one more time for my students. The exam results helped me to understand this. So, of course, we use [assessment] to reflect on our teaching and students' learning. (Buket, 13)

Buket is a good reflector as she makes changes in her teaching based on the exam results of her students. Apart from that, Ekin mentions how they make use of assessment results institutionally, together with her individual practices as a teacher. She underlines that the people in her institution's testing and curriculum units reflect on the assessment results for the betterment of those results, along with instructors' own reflections.

[...] In my institution, we try to understand where students are struggling and how we can treat these problems. This is how we try to use the students' results. I mean, as the [testing] unit. As a teacher, I can try to see what my students need and meet these needs by providing them with extra readings, for example, or teaching them more strategies about reading and reteaching some grammar again. (Ekin, 12)

[...] When it comes to exams, we always give feedback after the exams; we answer every question together with our students. [...] (Ekin, 12)

Additionally, Ebru mentions how she makes use of assessment results to improve student learning. She uses her office hours to provide individual feedback to her students based on the assessment results and her observations.

[...] I have been teaching English for... How many years now? Twelve, I guess. Right now, I feel like when I communicate with a student, even in the same classroom, there are different levels. I mean, each student is strong at one skill and not good at another skill, so I can easily detect which student needs what kind of extra activity to improve himself/herself. So, I use my office hours. I call the students and say something like, "You need this kind of study, so please focus on these extra sources." or "You need listening, you need writing." or "You need to do some grammar exercises." (Ebru, I2)

Nonetheless, even though teachers try their best to get help from assessment results to improve student learning, they indicate some concerns regarding this topic. One of the concerns is about the time constraints teachers face. They underline that the necessity to go on with the curriculum or time limitations make it difficult to allocate more time to feedback or revision of the necessary topics. Elif is the first participant addressing this issue. Her ideas about making use of assessment results are given below.

[...] Of course, we check the results to see if [students] did good on listening, and if they did bad on reading. Then, we should do more reading exercises or say, "Let us revise this topic over the next week." We try to help them. Maybe they can want this already. They can demand, as I said, if they are reflective enough, they can demand to study or focus more on ABC. But I think we do not have enough time. We cannot stop after an assessment for a week, for example, and say, "Okay, you have this, this, and this missing, so let us focus on these for some time. Then we can move on." That is the ideal, right? But we do not have a chance; no institution has that chance. We need extra time with students, and this is discouraging for both students and teachers. (Elif, I1)

The others who participated in the first interview session agreed upon this point. Along with Elif, participant Esra recommends that her students study individually after the lessons due to the time constraints she encounters.

I agree with my colleague Elif, and sometimes I advise students and say, "Okay, you need to focus on this topic. Study more on this topic, please, after the lesson because you have a lot of mistakes." After a writing task, I give such feedback. But yes, we do not have enough time to focus on what they need because we need to go on, and we cannot go through what they need deeply. (Esra, I1)

Overall, the participants point out that they make use of assessment results either to make changes in their teaching or guide their students through some suggestions to make them study individually due to different reasons, such as time limitations. However, it can be said that EFL instructors in this study perceive assessment as a way to improve the language learning process of their students through many different assessment tools and feedback, which leads to continuous revision in the teaching, implying the idea of AfL. Furthermore, instructors hold themselves accountable by reflecting on their teaching and making revisions when they are required to improve their teaching, as also discussed by Zaimoğlu (2013).

4.1.1.4 Meticulous Design

While explaining the improvement conception, Brown (2008) points out that assessment should be reliable and valid to represent the actual knowledge and performance of students since assessment results have to provide constructive feedback. Elif and Leyla also highlighted this point. It can be understood that even though their years of experience in teaching are different, instructors have encountered and observed some assessment practices that were not designed carefully, which might make assessment meaningless. Thus, they underscore that assessment tools have to be designed attentively.

Well, assessment sometimes, if not carefully designed enough, we can just grade students, but they may not always get the necessary information or feedback from that assessment. It could be just for the sake of testing somebody. And it is difficult to prepare, of course, a “well” assessment, let us say. It should be created and thought carefully while designing. Multiple-choice tests, for example, cannot always serve their purpose. (Elif, I1)

I was also thinking about a very specific example about writing, and as Elif said, if it is not very carefully designed, writing tasks, exams, or anything else do not serve much. (Leyla, I1)

Ekin, who has been working in the testing office in her institution, also implies the point of being meticulous while designing assessment practices since they need to be helpful for the learners in the long run.

[...] The important thing here is that: do your assessment methods align with your curriculum or do you try to make your curriculum align with your testing method? I mean, if it is the first one; if your assessment methods are based on your current curriculum, it means you are on the right track to help your student with the results of your assessment. (Ekin, I2)

In short, many of the nine participants repeatedly indicated the importance of assessment for themselves by exemplifying how they got help from assessment results to improve their students' language skills. Yet, it should be noted once more that teachers may hold different conceptions simultaneously, and they can even have conflicting ones together (Lutovac & Flores, 2022). Regarding this, participants' remarks that can be situated on the irrelevance conception will be presented in the next section below as the irrelevance conception had the second highest mean score.

4.1.2 Irrelevance Conception

The second most agreed conception by the participants in this study is the irrelevance conception, as suggested by descriptive statistics results ($M=30.13$). Surprisingly, many other assessment conception studies argued that irrelevance was among the least agreed conceptions. For instance, Brown (2002) found that irrelevance conception was the least agreed upon alongside student accountability conception in his study and stated a negative correlation

between improvement and irrelevance conceptions, which is not the case in this study. However, Remesal (2011) indicated that teachers may hold various and conflicting conceptions about the effects of assessment on teaching and learning. Likewise, in this study, the participants also have conflicting conceptions concurrently. This may imply that even though EFL instructors are aware of the importance of assessment as well as their crucial role as assessors, some factors, such as traditional high-stakes assessment perception in Türkiye, lack of knowledge, low-quality exams applied on an institutional or nationwide base and others result in a negative perception towards assessment, which is also indicated by Azis (2015). He discussed that teachers in his/her study held conflicting conceptions since their willingness to use process-oriented assessment practices for improvement was hindered by the national exam policy. Furthermore, the gap between instructors' ideal assessment context and the real assessment context in their institutions may lead them to have different and opposite conceptions because sometimes, although not all the time, the pre-defined assessment practices force the participants of this study to follow the practices that they do not prefer much. Some other underlying factors were highlighted while explaining the irrelevance conception in the literature review chapter based on Brown's (2008) explanations. These factors can range from the belief that teachers' observations are sufficient to make decisions about students to associating assessment with testing and thinking that it causes negative emotions, such as stress and anxiety. Furthermore, standardisation problems and not making use of assessment results for improvement purposes may make assessment irrelevant for some people. Finally, exams might be considered inaccurate indicators of students' actual knowledge, which is a point included in the literature review chapter. In light of these, focus group interview sessions could open a door to understanding the underlying reasons for the conception of irrelevance, as presented below.

4.1.2.1 Exams as Inaccurate Indicators of Student Performance

During the interviews, some factors that might affect students' exam results were mentioned, which makes it inevitable to include different assessment practices in the learning process to draw a correct conclusion about students' performance and knowledge. Firstly, as presented above, Elif underlines the importance of designing assessment methods and preparing exam questions carefully. Otherwise, exams may not reflect the correct information about students' knowledge.

Well, assessment sometimes, if not carefully designed enough, we can just grade students, but they may not always get the necessary information or feedback from that assessment. It could be just for the sake of testing somebody. And it is difficult to prepare, of course, a “well” assessment, let us say. It should be created and thought carefully while designing. Multiple-choice tests, for example, cannot always serve their purpose. (Elif, I1)

In another part of her speech, Elif also mentions that there might be some other underlying factors for students' failure in an exam rather than their knowledge of the topic. She illustrates this with her own experiences. She believes that the sole usage of exams might result in wrong interpretations of students. This makes it necessary to add other components to assessment practices. For instance, Elif highlights the point of observing students all the time.

In the department, I teach ESP [i.e. English for Specific Purposes] and those students already know some level of English. I do not teach them English, but I try to give them English in a different context. When I look at their exam results, most of these students fail because I know that they do not study for the exam. That does not show they do not know these topics. It shows that they are lazy. It does not show that I could not teach or that they did not learn, but just because of laziness, they do not remember some words. Maybe they misunderstood the instructions on the paper, and so on. So, I especially assess students during the lessons. I guess over time, during lessons, I can understand if they learn or not. Of course, exams or formal assessments are necessary, but I think a teacher cannot judge their students or their learning just by looking at those grades, mistakes or correct answers. (Elif, I1)

Looking at her statements, Buket seems to have a shared viewpoint. She builds upon the topic with the example of her own story as a student. She mentions entering Middle East Technical University (METU) with a great score, even though she could not speak English properly. She concludes her example by stating that exams are insufficient to conclude students' knowledge. Later on during the interview, she expresses her opinions about exams explicitly.

When I first entered METU, I had 99 correct answers out of 100, and I was not able to speak English. My instructor asked me to introduce myself, and I said I could not speak English, and I thought [it was] okay. As a picture, I was a perfect student, but when I wanted to speak English, I could not. So, if we come back to your question, I think just asking questions on a test does not assess what the students can achieve in real life. [...] (Buket, I3)

[...] I think the exams that we give the students these days are kind of decontextualised, and they do not give us the whole picture of what a student can do. [...] I think just asking questions on a test does not assess what the students can achieve in real life, but if you add other components to this, which I think I can manage, [it is better.] In my humble opinion, I can do that. (Buket, I3)

Hatice's perception follows a similar direction. Firstly, she points out the possibility of students cheating in exams, which might affect the exam results and lead to incorrect conclusions about students' performances. Secondly, she mentions her experience of designing and preparing an exam for her students for the first time. Again, she does not feel fully competent while creating exams, and she is doubtful about the results that she gets from these exams. Thus, she points out that exams should be supported by a variety of other assessment practices, such as in-class assessments that involve peer or self-assessment methods.

[...] I was stuck because it was my first experience preparing for a midterm exam. Okay, I kind of prepared a written test, but while I was preparing that test, I did not even believe that [it was] okay. It is not the only way to see my students' level. They can cheat in exams even if they do not know [the topic]. So, I guess the follow-up strategy for assessment should be changed in some way. Maybe

not as an instructor, but it is a departmental issue because sometimes, we can just apply the assessment strategy that the school needs. [...] (Hatice, I3)

Moreover, in the excerpt below, Hatice underlines the importance of developing a profound understanding of students' needs as well as taking their emotions, such as exam anxiety, into consideration since these are effective factors to be considered in the assessment process.

[...] I should also understand learners' needs because my thesis' focus was learners' anxiety and their readiness to learn the English language. Therefore, their readiness to learn English or study English affects their anxiety level to some extent, and instructors should be aware of this situation because some students might not be ready for these assessments, so [teachers] should take notes for these students. Otherwise, we just assess, and they cannot reflect their real performance on the test, so it should not be the only way. We should also see their personal needs. [...] (Hatice, I3)

Finally, Songül shares her experiences and observations on the day of the interview. She believes that tests and scores are sources of stress for students. This stress may make them focus on getting a good grade by changing the focus from learning. The example she gives at the end of her speech demonstrates how students just try to memorise phrases or sentences due to their focus on grades, rather than the learning process itself. Even when students seem successful in exams due to memorisation, whether they have achieved the intended outcomes or not is questionable, which is parallel with the PSTs' conceptions of assessment in the context of Xu and He (2019). They pointed out that PSTs were inclined to think of memorisation and imitation in assessment, whereas this inclination evolved into comprehension, problem-solving, and creativity in in-service teachers.

[...] Today, we had a midterm, and I was in the classroom ten or fifteen minutes before the exam. [...] I saw that students were trying to memorise certain phrases or structures five minutes before they were going to take the test or whatever. [...] I read the papers today before I joined [this session], and all of them got it wrong. There is no actual photographic memory of what they are doing because they do not know; they do not know articles, or they do not understand. They just hear them, see them, and try to copy them. [...] Also, I see that they are writing the same thing: "The researchers have shown that eight out of ten people have been playing online video games." This cannot be correct; [there is] no way. [...] (Songül, I3)

What Songül illustrates above relates to how assessment is perceived in the Turkish context. Hence, more examples concerning national contextual factors are provided in the following section.

4.1.2.2 National Contextual Factors

As indicated by Tsagari and Vogt (2017), national authorities strongly influence the assessment practices of teachers. During the interview sessions, many participants overtly uttered that people in Türkiye are used to traditional assessment methods, such as exams, which take place in their institutions continuously, even though they do not favour them. Similarly, some instructors shared their own stories when they were students to illustrate the effect of exams on

their perceptions of success. They indicate that their understanding of success was in line with their exam scores. Now that they have been working as EFL instructors for some time, their observations demonstrate that students in today's world also tend to prioritise traditional methods, although the instructors now pay more attention to performance-based assessment than these traditional forms. For instance, Ebru states that she attempts to convince her students about the importance of portfolio assessment though she is not taken seriously by her students. Clearly, it is seen that even though the participants of this study do not indicate that assessment is irrelevant, they highlight the negative effect of national contextual factors.

Initially, Ebru has been teaching for around twelve years. She mentioned that exam scores were given so much importance when she was a student. She thinks that this idea is still valid for students today, which might be because of the perspective of assessment that people in Türkiye are used to as a nation. Concerning the preparatory school context, students also have to pass a proficiency exam to be able to start the first grade of their departmental courses. As Güngör and Güngör (2021) indicate, universities are autonomous in the decision-making process of these high-stakes proficiency exams; however, they are quite influential in educational practices. Thus, exams and exam scores are considered to be of paramount importance. In this vein, Ebru's following statements demonstrate students' tendency to think that exam results are crucial rather than their progress.

When I was a student like a high school student, I thought, "If I have a good score, I am a good student." That was the only association that I had. [...] Then, I realised that it was not the only thing. Right now, I am trying to convey this message to my students, and I do not know how to articulate that because their mindset is similar to what I had when I was a high schooler. So, right now, I am trying to explain to them how portfolio assessment is important and why. Well, I try to do my best, but at the same time, some of them take it seriously, and some of them do not. [...] (Ebru, I2)

Melda underscores the impact of standardised tests in the Turkish context. She expresses that she favours other types of assessment practices, such as peer assessment or self-assessment methods even though her institution follows a different policy, which she relates to the way assessment has been perceived in Türkiye. Schellekens et al. (2021) pointed out a similar perspective that teachers felt a clash between two types of assessment since graded assessment hindered assessment for learning purposes. In this respect, it is worth noting that when assessment practices are not real reflections of what teachers believe, their level of acceptance will be low, as pointed out by Brown (2008).

What I would rather we did in our institution is to have fewer exams and more classroom assessments, but I think this is because of the country-level context and because we are so used to standardised tests. We do have midterms, and we talked about them, but I think they are too frequent. [...] (Melda, I2)

Moreover, during the latter part of the meeting, Melda indicates the effects of contextual factors on students' perspectives. Melda's observations suggest that students are inclined towards test scores or teacher-centred assessment practices.

In our Turkish context, assessment is generally seen as something done by the professor or by the teacher, but we do have peer assessment and self-assessment, right? So, we could look at how assessment impacts their learning, maybe from this perspective as well. Generally, we do have all those things. If something is not in the test, students generally do not study and pay attention to it. Also, generally, this is what I have seen; assessment given by other peers does not generally affect their learning. Maybe they do not trust their peers. Sometimes, they do not trust themselves, and they may say, "How should I evaluate my own learning?" From that perspective, they pay more attention to the feedback or the score coming from the professor or the centralised test. (Melda, I2)

Similar findings can be seen in some other contexts as well. For example, a research study in Hong Kong by Brown, Kennedy, Fok, Chan, and Yu (2009) found that the high-stakes exam context in Hong Kong led teachers to focus on exams, which is similar to the Turkish context to some extent because of the focus on high-stakes exams. Moreover, Buket's example presented below, which demonstrates her perspective of assessment as a student, is similar to Ebru's case mentioned above and is considered in the framework of the national contextual factors. She, as a student, considered herself a successful language learner based on her exam results, which did not assess speaking or listening skills.

When I first entered METU, I had 99 correct answers out of 100, and I was not able to speak English. My instructor asked me to introduce myself, and I said I could not speak English, and I thought [it was] okay. As a picture, I was a perfect student, but when I wanted to speak English, I could not. [...] (Buket, I3)

Based on the examples of Buket and Ebru, both this study and the study of Xu and He (2019) indicate that conceptions of assessment can alter over time from an exam-oriented view towards a process-oriented view. Besides, elsewhere in her speech, Buket argues that specific assessment ways are applied in the Turkish context; however, she addresses the need for some nationwide changes by illustrating this with spoken assessment.

[...] I think we lack assessing interaction in most preparatory schools, not only in our school, because we think this is very difficult. [...] So, an addition of paired assessment, group assessment, and assessing the interaction is something that needs to be added in most of the preparatory schools in Türkiye, I think. (Buket, I3)

When it comes to Leyla, even though she does not explicitly express being used to traditional methods of assessment, the excerpt below implies that students do not care about feedback, which may be directly related to their focusing on exam results, not the learning process itself.

I think the only assessment that I reflect on is the writing tasks because I also spend a lot of time on it by marking and writing feedback. So, I like just as well to spare some time and give feedback. But [...] students do not want to get feedback most of the time. [...] (Leyla, I1)

Last but not least, Elif uses similar words and points to national contextual factors. Also, while reflecting on her own classroom assessment practices, she confesses that she also adopts traditional methods.

[...] When it comes to vocabulary, reading and listening skills, structure, etc., I guess we are used to those kinds of exams and assessment types as a nation. We know how to prepare, or we know how to assess or prepare our students for that kind of assessment. (Elif, I1)

[...] When I think of myself individually, how do I assess students in the class? I could not find a lot of things. I go very traditionally, I guess. [...] (Elif, I1)

Overall, the participants of this study have a high level of awareness concerning national contextual factors, and some of them exemplify how their perspectives as students have changed through their teaching careers. Even though the perception of the participants has changed, these examples and explanations mostly show the tendency of students and people in the Turkish context to consider students as successful based on their exam scores and provide a framework for the ongoing assessment policies in Türkiye. Moreover, the participants were asked whether they feel competent regarding assessment or not since they have been educated in this context for all their lives. Their answers are presented in the next section.

4.1.2.3 Teacher Competencies

While defining the irrelevance conception, it was included that insufficient training of teachers might be related to assessments' being comprehended as irrelevant (Stiggins, 1997, as cited in Vardar, 2010). It should again be highlighted that the participants in this study do not explicitly point out that assessment is irrelevant or should be omitted; however, the analysis of focus group interviews revealed that many participants did not find themselves completely competent considering their knowledge of assessment and assessment practices. Moreover, the insufficient pre-service training provided to instructors and the failure of assessment courses in shaping their perceptions of assessment were underscored by the participants. In this vein, it is worth noting that the insufficient training or the incorrect perception that these courses gave might have led the participants to perceive assessment as irrelevant while choosing items on the TCoA-III.A.

Firstly, while talking about their competencies concerning assessment, the participants indicated that they felt insufficient; however, they focused on different points. Besides, it should be kept in mind that four participants have worked/have been working in the testing office of their institutions. (i.e. Elif-I1, Ekin-I2, Melda-I2, and Buket-I3) and those instructors feel more comfortable than others concerning assessment, such as preparing exam papers. Nevertheless, they also feel a need to improve themselves. Observations and opinions of the researcher concerning the topic are written in her self-reflection journal.

What I observed based on the participants' facial expressions also indicated that they definitely needed more knowledge when it came to assessment. However, what they really need is practical

information about real life, not theories or names, such as formative summative assessment etc. Even though some participants stated that they are familiar with preparing exams etc., they feel alone in the classroom, and they mainly rely on their own observations. What I think is arranging some seminars in relation to real practices as well as cooperation between teachers would be better because while talking, I guess, the participants also realised that they have some common issues, so not feeling alone can make them work cooperatively and find some solutions together. (Journal entry, December 13, 2023)

Below are the reflections of instructors on their assessment competencies. Firstly, Elif indicates that she has prepared exams as a part of her job in the testing office, making her feel more knowledgeable compared to other participants. However, she does not feel competent enough to make in-class assessment-related decisions.

I think all teachers, first, should be more or less knowledgeable about types of assessment. [For example,] what is important while assessing students, what are the assessment types, and what are the important things we should know about it? But I think we do not focus on that area much as teachers. I feel a bit more confident than you as a result of my duties. I worked as a tester and prepared different assessment types, but still, those were just formal assessments: tests, multiple-choice tests, and writing exams. [I mean] the very mechanic part of the assessment. Still in class, while assessing students, I always think, "Okay, I can prepare those students for the exams very well [...], but when I think about their writing or speaking skills, I still do not feel that confident." Writing and speaking are already difficult skills, and [students] avoid speaking during class. So, I always try to make them speak, but I am not very aware of how to assess when they speak and how to give feedback, for example. (Elif, I1)

In a similar vein, Melda has been involved in teaching for more than twenty years and has also been working in the testing office for quite a long time. When it comes to her competencies, she addresses the need to get extra training if she has to change her current role and prepare assessments for students at the beginner level.

I have been involved in assessment, but my experience is mostly in proficiency level assessment, so we do have a level that is about CEFR level B2. This means that I am used to that level, and I can assess reading, vocabulary, and grammar at that level, but what I am not feeling so competent is an early-level assessment at the beginner level. [...] What could I do for that if I had such a role in my institution? I would probably need training and maybe some self-study. (Melda, I2)

Buket also had to be a part of the testing office in her institution, which she did not enjoy much. Nevertheless, she believes that she is a good assessor of students' language skills. While explaining that, she draws a good picture of her understanding of assessment, which is in line with AfL.

I do not have much experience with [testing]. I had just a little and I did not like it, but I can say that I am experienced in a way that I can assess my students' language competence. But it is not only grammar, it is not only vocabulary, it is not only listening; all together. I can assess my students, but it does not mean that I am giving them exams or things like that. I can do this as an ongoing process [...] So if we come back to your question, I think just asking questions on a test does not assess what the students can achieve in real life, but if you add other components to this, which I think I can manage, [it is better.] In my humble opinion, I can do that. (Buket, I3)

Ebru shares her opinions about her competency regarding assessment. She is a good observer and understands her students' needs and skills. Still, she does not feel confident in assessing students in a standardised way.

[...] I have been teaching English for... How many years now? Twelve, I guess. Right now, I feel like when I communicate with a student, even in the same classroom, there are different levels. I mean, each student is strong at one skill and not good at another skill, so I can easily detect which student needs what kind of extra activity to improve himself/herself. So, I use my office hours. I call the students and say something like, "You need this kind of study, so please focus on these extra sources." or "You need listening, you need writing." or "You need to do some grammar exercises." (Ebru, I2)

[...] I mean, [the assessment courses that I took], of course, gave me some information about what to do and what not to do, what to pay attention to, and what to do in the classroom. But at the same time, I feel like I am not competent to test a student in a standardised way. (Ebru, I2)

Apart from that, it is apparent that Esra and Ekin have the same opinion since they indicate the necessity of improving themselves concerning assessment. Ekin, who works in the testing office at her school, focuses on preparing exam questions. On the other hand, Esra approaches the topic from a broader perspective.

[...] I am not good at writing distractors. This is a confession; this is sad, but it is what it is. I am not good at writing distractors, especially in reading. (Ekin, I2)

I feel that I need to learn effective ways of assessing the language skills of our students. (Esra, I1)

The other participants also elaborate on different elements of assessment. Leyla believes that she does not feel very confident while tackling students' mistakes and providing feedback in the classroom.

When it comes to regular speaking or writing exams and marking, to be honest, when you look at a rubric and when you have done marking a bunch of times, you get used to it. There is a certain expectation of what the students know and where to deduct points. But when it comes to in-class assessments, I, too, feel very confused. On the one hand, I am very happy that students are producing the language, and I do not want to discourage them and be like, "Okay, the teacher fixes all the mistakes, so I am just not going to speak so that I do not make mistakes." However, within class time, I see that they go through some process of fossilising that incorrect language, so I want to fix that, too. So, I do not feel okay about my skills or in-class assessment. (Leyla, I1)

Hatice, on the other hand, is more focused on the difficulty of preparing exams for her students as she has experienced that for the first time. However, looking at her speech, it is evident that assessment has many different elements, and she feels more comfortable with AfL tools.

[...] I felt insufficient in preparing a mechanic test because it was not easy to adapt all the subjects since I taught different classes. After all, I sometimes did not teach [some topics to some classes], and if I had not [included] this subject [in the exam], the other classes could not have been assessed. But if I had included that subject, the students who were behind the curriculum would have felt insufficient. It was difficult for me to prepare for a written exam, but it might be different while preparing a spoken assessment, role-play assessment, presentation, or portfolio assessment. [These] could be different, maybe. (Hatice, I3)

Lastly, Songül believes that she is knowledgeable about the topic, and she tries to improve herself by reading, searching, and asking questions to the people in charge.

I think I am okay at assessment, but it is only because I did research on different types of assessment. Also, we have rubrics that are prepared by the testing office that we use in assessment, and I try to read them and ask questions if there are some question marks for me. [...] (Songül, I3)

All in all, most of the instructors taking part in this study indicate that they need to improve themselves in terms of the application of some assessment practices. Also, it is a field that requires effort and training regardless of their years of experience in teaching.

4.1.2.4 Ineffectiveness of Pre-Service Assessment Courses

The code “teacher competencies” mainly indicated that instructors needed extra training to improve themselves in assessment, which might be linked to their ineffective training during their pre-service education. Hence, the participants’ opinions about their pre-service assessment courses are included in this section. To begin with, Elif directly commented that these courses did not prepare her well for real teaching contexts. She uses the following words concerning the course “Language Testing & Evaluation.”

[...] Well, I always criticised my university's approach. I learned a lot of things, wrote a lot of research papers, read a lot of articles, etc. [I remember] all the terminology and methodology about teaching, but it is not about real life. The moment I graduated and started working, real life was not like it was given in the books. I always criticise this aspect of my university. But I remember what validity and reliability are about assessment or assessment types: formal assessment, summative assessment... I can list all of them, but in practise, I mean, how to assess during class while a student is speaking, how to assess if he/she is doing well, what aspects should I note down or should I ignore? Nobody told me anything about those. (Elif, I1)

In another part of her speech, Elif again states that the practice part is neglected in many pre-service courses designed for PSTs.

[...] I think those courses, Language and Assessment, Language Testing, etc., are useful if you want to become a testing member because when I became one, I saw, “Okay, now it makes sense. I prepare a test and now I know what validity and reliability are for an exam.” But, as I told you before, assessment is different during class, and I think something is missing in the education faculties’ curriculum in that aspect. I am an ELT graduate, but I guess that practice in each aspect of teaching is missing, not only for assessment. (Elif, I1)

Melda, one of the most experienced participants, indicated that extra training was necessary for her to improve herself when she graduated and started working as a teacher. Also, she improved herself through her experiences over time, similar to the statements of Mertler (2003) who concludes that teachers’ practical experiences teach them more about assessment in comparison to pre-service education.

Twenty years ago, when I was a student, there was a testing course, and we did have a nice instructor. But when I first graduated, I saw that it was not enough. And whatever I learned in the field, I learned on the job. Plus, I had to receive in-service training and a lot of other outside training. In a nutshell, the language testing course that I took was not enough. (Melda, I2)

Besides, Ebru and Buket complain about the same issue. First, Ebru argues that she does not feel competent in assessing her students in a standardised way, although she took courses on assessment. Correspondingly, Buket pointed out that preparing tests was a topic that she did not

feel competent about in the early years of her teaching experiences. She believes that some topics should have been covered more comprehensively in the assessment courses of PSTs.

During both my bachelor's and master's degrees, I took language assessment courses. But in the end, they, of course, gave me some information about what to do and what not to do, what to pay attention to, and what to do in the classroom. But at the same time, I feel like I am not competent to test a student in a standardised way. (Ebru, I2)

Especially about writing an exam, I think we need to be given some, I do not know... I remember from my own experience. I did not even know that while writing a multiple-choice question, you cannot write, for example, a grammatical mistake there. Let us say you are assessing the simple past tense, and you write "haved" in the choices. I thought this was okay back then, but I learned that it was not and thought, "I should have learned this before." So maybe even if it is not the kind of course, writing questions, analysing, and discussing them together before you do these kinds of things, I think it is necessary. What is included in the syllabus of [assessment-related] courses can be changed, maybe. (Buket, I3)

As previously mentioned, the study also included participants who graduated from the ELL and Translation & Interpretation Departments. When it comes to them, they reflect on their competencies and how they have improved themselves concerning assessment. Specifically, two instructors graduated from the ELL Department, while the other one graduated from the Translation & Interpretation Department. This means that these instructors did not take any formal courses about assessment during their pre-service education. However, Esra states that she attended a teaching certificate programme (i.e. the Pedagogical Formation Certificate Programme), and she reflects on the courses she took there.

I agree with my colleague Elif. [...] Actually, I have [a teaching certificate] and I think it was a really well-developed programme. But I felt that our professors did not know what we did in our language classes, and they did not have that kind of experience, maybe. They knew the theories, but they did not have practical experiences. Therefore, they could not teach, maybe. They could not know what students needed in language classes. Therefore, something was missing in terms of giving the correct ways of assessment. It was nothing more than the traditional ones. (Esra, I1)

On the other hand, Ekin mentioned that she did not take any pre-service courses concerning assessment. To compensate for that, she participates in-service training and improves herself while working as a part of her duties. As seen below, she admits that she has a long way to go and improve herself concerning the topic.

I am not an ELT graduate, so I had not taken testing courses during my bachelor's years, but when I started to do testing, I started to get some in-service training. So, I learned the job both there and by doing it, and I am still learning and still have a lot to learn. (Ekin, I2)

Additionally, Songül is also a graduate of the ELL Department. She stated that she did not take any assessment courses either. However, during her speech, she points out that taking courses on the topic would be better, even though she does not expect them to be very comprehensive. Her idea of spending limited time on assessment contradicts the other opinions implying that one should continuously improve oneself in terms of assessment, or just one course on assessment is not sufficient.

I think yes [training on assessment is necessary], but at the same time, I do not think you need to have a specific course related to that at your undergraduate level or in your master's education, maybe. [...] I think you can do deep work for five weeks, and you will be done with the assessment part. Maybe a course incorporated with assessment courses, a major one with some parts dedicated to assessment, would be okay. I think it would be better than just having a 14-week or 15-week period of assessment courses. (Songül, I3)

Unlike Ebru and Buket, who mainly put an emphasis on exam preparation and testing part, Leyla underlines the necessity of focusing on in-class assessment. She is mainly concerned with the in-class assessment and feedback since she does not actively participate in the exam preparation step as she is not a member of the testing office in her institution. Her statements can be evaluated considering the words of Hatipoğlu and Erçetin (2014). They argued that assessment is generally conducted by the testing office, although the competency of the teachers working at the testing office is a point of discussion. Furthermore, leaving assessment only to a small number of people working at the testing office results in a decline in other teachers' competencies and knowledge as they are not encouraged to improve their skills. In the interview, Elif strongly supports Leyla's opinions after she stops explaining her opinions. Thus, it is essential to underscore that many instructors underline the importance of feedback, process learning and learner-centred methods, especially when students complete their writing tasks. In this regard, the opinions of the students studying in higher education point to the same spot in Flores, Simão, Barros, and Pereira's study (2015). They stated that the ones who address learner-centred assessment methods conceived assessment as more effective and fairer (Flores et al., 2015).

Personally, I think of certain parts of testing courses in which we learn about how to write questions, etc. That part was necessary only if you planned to be a testing member because during my university years, I thought, "Okay, this is going to come in handy." But when you go to work, you see that there is a whole unit for this and you are not going even to see the questions before the exam, so that was very unnecessary. I would have appreciated the course purely focused on how to give feedback. I think that would be much, much better because I learned that skill at work. [...] (Leyla, I1)

Lastly, rather than pre-service courses, Hatice mainly mentions the courses she took during her master's education. This may imply that she does not remember her pre-service assessment courses and might be interpreted as the fact that these courses were not influential in her understanding of assessment.

I am trying to remember, but I do not know. We did not have a specific lesson for assessment, but while I was doing my master's degree in the UK, there were lessons like the written course lesson. We also learned some assessments but in a written way. It was like a student sample presentation, student sample portfolios, preparing a questionnaire, gathering the data, and assessing them. So, how do the students on campus have an idea for a specific point? I remember that we analysed and tried to assess the student questionnaires, but it was a minimal assessment, like an Excel sheet, and it was not too detailed. [...] (Hatice, I3)

In short, it is deduced that the participants of this study did not believe that they were fully competent in assessment even after they graduated and started working, which inevitably led

them to improve themselves by engaging in extra training, as in line with the findings of Tsagari and Vogt (2017) who discussed that education on assessment was neglected while preparing PSTs for real-life in the context of their study. Therefore, the courses offered to PSTs should be questioned since the main aim of assessment course/s is to provide them with the necessary skills concerning assessment (Kyttälä et al., 2022).

4.1.2.5 Assessment Courses in Shaping PSTs' Mindset

During the coding stage, the researcher coded this part differently. However, after rereading the statements of the participants, the name of the code changed to the last form given above. In addition to the statements presented above, the instructors were requested to reflect on how assessment courses shaped their understanding of assessment. Elif and Melda do not think their pre-service assessment courses gave them the right impression about assessment. Specifically, Melda states that one course is insufficient to shape PSTs' understanding of assessment. In addition, she underlines that assessment is a field that requires continuous training to improve oneself and keep up with the changes in the field, as also indicated by the participants of Levi and Inbar-Lourie (2020).

I remember when I was taking those courses, assessment scared me. I thought writing questions and those concepts were difficult and I thought, "No, I cannot do that." Maybe because of the course design or the instructor... It was a long time ago, but I did not like testing at all, but look where I am now. I mean, just three years after I graduated, I became the curriculum and testing member, and I dealt with testing and assessment for how many years... and I liked it when I started doing that. So, I will refer back to the previous question, but the courses were theoretical and did not give the right impression about assessment. (Elif, I1)

[...] That year in your teacher education, which was my third year, I think... I guess it was a bit early because I did not understand what the test was and what I was doing. We did all sorts of stuff; we wrote tests, provided feedback, did item analysis, and so on. However, when I graduated, assessment was not something like that. There should be a rationale behind it; there is an administration, exams are controlling you, and everything has to be taken into consideration. So as far as I think, one course is not enough. But assessment is something just like teaching, maybe; you should feed it constantly and get training. [...] (Melda, I2)

Ebru and Buket, on the other hand, start referring to the courses they took during their master's education. Thus, it can be thought that the influence of master's degree courses cannot be discarded, whereas this conclusion cannot be directly drawn for pre-service courses.

When I was a student like a high school student, I thought, "If I have a good score, I am a good student." That was the only association that I had. [...] Then, I realised that it was not the only thing. Right now, I am trying to convey this message to my students, and I do not know how to articulate that because their mindset is similar to what I had when I was a high schooler. So, right now, I am trying to explain to them how portfolio assessment is important and why. Well, I try to do my best, but at the same time, some of them take it seriously, and some of them do not. So yes, these courses positively shaped my understanding, I can say. (Ebru, I2)

Buket indicates that she has gained a holistic perspective on language assessment through her master's degree courses.

I think this happened to me in my master's degree. I realised that assessment is not only "Okay, now I am assessing your grammar, or I am assessing your vocabulary." In my master's degree, when I was taking one course, I realised that we have to have a holistic view of assessment and look at the students' outputs in that sense. That is why, for example, [...] when I look at the rubric in spoken assessments and see grammar, vocabulary, etc., I get furious because spoken assessment is not only that. It is about being able to interact. It is about the interaction competence. It is beyond that, and the course that I took in my master's degree helped me gain this understanding. (Buket, I3)

Reflecting on her courses in the teaching certificate programme, Esra believes that her opinions about teaching and assessment have changed after she started working, which might be interpreted as not gaining a fully correct understanding concerning assessment when she was a pre-service teacher.

I was like my colleague Elif when I was a student because I thought it was a really hard job, and I thought that some people with magical powers wrote those coursebooks and other things. Then, five years later, I wrote a workbook, for example. But yes, [...] I thought it was a very hard thing to do, but later, I got used to that when I started teaching, and I understood how I could assess students or evaluate what they knew. (Esra, I1)

Just like Esra, Leyla mentions that her perception of assessment has shifted. The assessment course/s focused on the mechanical part, which is mainly about exam preparation. However, she believes that assessment is not always boring, challenging, and mechanical.

I was also very scared, and for some reason, I thought that [testing] was a very boring and mechanical job and that I had to be very careful about teaching certain items in class so that students could do it in the exam, but I thought it was too mechanical. Of course, there are certain parts that have to be mechanical, strict and clearly designed. However, when it comes to ungraded assessment, for example, it does not have to be very mechanical; it can be creative. (Leyla, I1)

Hatice, in response to this question or other questions during the interview, was primarily concerned with students' individual needs, which she started prioritising when she was a master's student; however, she mentioned that she did not remember her pre-service assessment courses during the interview.

[...] The lessons that I took during my master's degree helped me understand that. I believe that the only way of assessing students is not the written way or making them scared of assessment. "Okay, I will grade you, and you will be assessed." [I mean] the idea of these things. So, the lessons that I took showed me that I should also understand learners' needs because my thesis' focus was learners' anxiety and their readiness to learn the English language. Therefore, their readiness to learn English or study English affects their anxiety level to some extent, and instructors should be aware of this situation. [...] (Hatice, I3)

This section reveals that the pre-service courses of the participants were ineffective in depicting a true picture of assessment in real life. Even though the participants do not imply finding assessment irrelevant, this negative impression that these courses conveyed led many participants to be distant or scared of assessment, even as teachers.

4.1.2.6 Standardisation Problems

While explaining the irrelevance conception, Brown (2002) points out that students believe that their grades depend on the instructors' own interpretation and judgement. This idea leads them to bargain to get fair results. Furthermore, he indicates that subjectivity is involved when an examiner assesses a student, which may negatively influence the meaning of the assessment (Brown, 2008). During the first interview, Elif drew attention to this standardisation problem although she did not imply that they find assessment fully irrelevant because of standardisation problems, which is agreed upon by other participants in the same session. As also underlined by Brown (2008), the researcher thought that this concern could be a code under irrelevance conception.

In interview one, subjectivity in grading is mentioned by Elif and the other participants seem to have a shared point of view. If this drawback is brought up in other interview sessions, it can be placed under irrelevance conception based on the explanations of Brown. (Journal entry, December 14, 2023)

[...] When there is a writing and speaking assessment, there are the assessors, of course, and there is not just one person assessing all the students, so standardisation is always a problem, I think. No matter how we study the subject and have those meetings to grade those papers or the speech, there could be a weakness in that sense. I mean, if somebody comes and examines our results, I can say that this could be our weakness. I think it is a general issue, not for our institution. The moment writing or speaking is involved and assessed by someone, of course, there are discrepancies. (Elif, I1)

When she stops talking about this issue, Esra agrees with Elif. Moreover, in the third interview session, Hatice highlights the same point.

[...] I guess the results of the proficiency exam do not reflect the students' real level because, as we know, there are some parts that assess students' level [in this exam]: writing, speaking, and mechanical parts. The mechanic test might be objective, but the written and the spoken assessments are more likely to be instructors' way of assessing. At that point, I guess there must be a standard for teachers' perspective because I might think of giving a score of 15 for a writing paper, but the same paper can be assessed as 25 [by another instructor.] That is a huge gap between the two instructors, so at that point, there must be a standard, and I do not believe that there is a perfect standardisation for the assessment of writing or spoken exams. [...] (Hatice, I3)

Obviously, instructors' experiences and observations lead them to think that standardisation is problematic while assessing students' language skills, regardless of the efforts to have a perfect standardisation. In this vein, Black, Hall, Martin, and Yates (1989) discuss that written standards should be discussed by the assessors to reach a shared comprehension. The problems mentioned in this study concerning standardisation were mainly concerned with the assessment of speaking and writing skills, which can depend on the subjective interpretation of the assessors. Therefore, the assessment results of speaking and writing skills can be considered unfair and might be interpreted in the framework of irrelevance conception. Thus, these concerns highlighted in the focus group interview sessions might have been the influential

factors leading instructors to choose items about irrelevance conception when responding to TCoA-III.A.

All in all, based on the findings above, it should be noted that the participants did not explicitly point out that they found assessment irrelevant; however, some points about irrelevance conception in line with Brown's explanations (2008) could still be understood from their statements. This means that although EFL instructors do not perceive assessment as totally irrelevant, there might be some points to be considered to make assessment more effective.

4.1.3 Student Accountability Conception

The third preference of participants is the student accountability conception ($M=12.64$), followed by school accountability ($M=11.55$), even though the difference between these two conceptions is not significant. Contrary to the findings of this study, the study conducted by Vardar (2010) revealed that student accountability conception had the highest value among the teachers teaching different subjects in 6th, 7th, and 8th grade, which Vardar associated with the competitive nature of Turkish education system full of high-stakes exams. The fact that student accountability conception is also agreed upon by the participants to some extent might be concerned with having a proficiency exam at the end of the preparatory school year. More precisely, both the participants of this study and the preparatory school students have been educated in a high-stakes exam country for all their lives by preparing for high-school entrance exams, university entrance exams, and others, which go on with the proficiency exams that they have to pass at the end of the preparatory school year. Thus, the proficiency exam will be the determinant of whether students will fail or pass based on a criterion and will have a substantial effect on students' lives. Therefore, this study argues that the educational system and its demands might influence teachers' conceptions, as also highlighted by Remesal (2011).

While explaining student accountability conception, Brown (2008) highlights that it is concerned with assessing students' performances with the help of criteria, which helps grade or choose them for higher levels of educational institutions by comparing them with their peers, and students are held accountable for their own performances and learning. When describing how she perceives assessment, Melda states that assessment is a complex term to define since it has many different aspects. Also, she directly implies student accountability conception with words such as certification or decision-making.

[...] As my colleague mentioned, we use all those things for just, I think, one purpose, which is to improve and contribute to student learning. But at the same time, we must do it for certification and

decision-making purposes, so, it is a concept that has many, many, many layers and aspects. (Melda, I2)

Moreover, as explained in the second chapter, making students accountable for their own learning may have some reasons behind it. For instance, students can be motivated by actively engaging in assessment. Additionally, it is a common practice to make students participate in assessment processes through peer assessment or self-assessment methods (Brown, 2008). Likewise, during the focus group interviews in this study, participants mentioned elements regarding student accountability conception, which are to be explained below.

4.1.3.1 Including Peer Assessment and Self-Assessment

Brown (2008) states that one of the reasons behind student accountability is student-centred classrooms that lead students to take responsibility for their learning. While talking about the assessment practices they use in their lessons, the participants mentioned that they try to include peer feedback or self-assessment by taking students' reactions and needs into consideration, unlike what is indicated by Tsagari and Vogt (2017). In other words, Tsagari and Vogt found out that their participants did not necessarily prefer alternative assessment tools despite their awareness of these practices. Nevertheless, in this study, all three instructors from the first interview with different years of experience mention that they include these kinds of AfL practices. Initially, Elif explains how she makes use of peer assessment in her lessons.

[...] Since I am not a main course teacher, I do not have any responsibility for doing writing or speaking tasks, so peer feedback for those is also not my duty. However, I use peer feedback very little when somebody answers something. I do not say, "correct" or "incorrect" but I receive feedback from the students. "Is it correct? What do you think? Let us correct our friend and give him/her the right version. Who wants to explain this?" Generally, I have question-answer sessions. [...] (Elif, I1)

While criticising the institutional practices, Melda indicates that peer assessment and self-assessment are practices that need to be more integrated into the curriculum since they meet the demands of the changing world.

What I would rather we did in our institution is to have fewer exams and more classroom assessments, but I think this is because of the country-level context and because we are so used to standardised tests. We do have midterms, and we talked about them, but I think they are too frequent. In my view, we could have fewer of them and more peer assessment and self-assessments, maybe more like reflection type, etc. We do have structured ones, but I believe they could be fewer in number, and [we could have] more peer assessment and working with others because the world is changing. In today's world, collaboration, communication, working with others, being a team member, and supporting and getting help from others are all really necessary skills. It is the same for language learning and assessment, which are transferable skills. If you can do it in Turkish, you can do it in English. If you can do it in English, you can do it in your workplace. Therefore, I believe we should shift our focus a little bit to other types of assessments. Structured and standard assessment is not enough in our world at the moment. (Melda, I2)

Concerning peer assessment or self-assessment, Ebru referred to her previous work experiences by highlighting the significance of self-assessment in the language learning process, which she associates with higher student motivation.

[...] When I first started teaching, I was teaching at a language school, and it was just a portfolio assessment. We did not have any written exams, and I felt confident about it because the students were there to improve their English. They had some realistic expectations and when we did the portfolio assessment, they were willing to participate, and they did not have a score anxiety, let me put it in that way. They wanted to learn English. [...] Because of that, portfolio assessment was really good, and they assessed their skills in some intervals. After one month, we went over all the skills, and the students were doing their self-assessment. We were discussing with them how well they were doing and what else they could do to improve their English. So, it was nice in that way. But when I started teaching at the preparatory schools, it was just like the written exam and back to high school, let us say. So, I sometimes question if our objective is to test the students or to improve students' English. Sometimes, they do not match well. That is why I feel kind of uncomfortable with our testing strategies, but I cannot come up with a working alternative. [...] (Ebru, I2)

Moreover, Esra and Leyla state that they prefer peer feedback when it comes to writing. They sometimes abstain from peer feedback since their students feel shy about providing feedback to their classmates under some circumstances.

In the class, I use peer feedback. For example, they write a paragraph, and I want the students to give feedback to their classmates. [...] I do not use it in speaking because they can be really shy and frustrated, so I do not use that. [...] (Esra, I1)

Especially when it comes to writing, [...] I do peer feedback within the class, but not always because it is very embarrassing for students at some levels, [...] so I try not to leave it to the students perhaps at the start, but towards the end, yes, I do peer feedback. (Leyla, I1)

Hatice, who participated in the third interview, indicates including self-assessment in her lessons. She requests her students to reflect on their knowledge and skills in intervals. By making them write notes without their names, she believes that students can share their sincere reflections, which, in turn, helps her decide the further steps that she should take as a teacher.

[...] At the end of my lessons, twice or once a week, I ask my students to have a blank paper and write what they learned or did not learn during these two weeks, and they anonymously write them. So, I can be aware of their situation. [I mean] whether they have learned or not. They feel less anxious when they do not write their names on these self-assessment forms, so I let them assess themselves: what they have learned and what they would like to learn in that subject before the midterm exam or the [final] exam. (Hatice, I3)

All in all, the participants mentioned that students should be accountable for their own learning, which manifested itself through the support of peer assessment and self-assessment methods. Their thoughts might be similar to the statements of Harlen and Crick (2003) who said that teacher assessments could be more motivating for students and less detrimental to their self-efficacy compared to high-stakes assessments. Furthermore, thinking that assessment can help students improve their learning is also related to student accountability conception, which is illustrated below.

4.1.3.2 Guide for Students

While explaining improvement conception, the code “assistance for teachers” was explained (See page 55). Even though the emphasis is less on the student side, instructors also think that assessment is helpful for students to improve their language skills by drawing a map for themselves. This perception is closely associated with the conception of improvement. In the first interview session, Elif and Esra mention how assessment can be helpful for students.

Learners also need to see what they have done and what they could not do on a task, for example, at the end of this assessment. It does not have to be graded or something but in any kind of assessment. That is like a sign for them or a guide for them to see which parts they have learned better or which parts they should improve. [...] They can reflect on their learning if they are clever enough, well, if they are good reflectors, let us say. (Elif, I1)

They can demand, as I said, if they are reflective enough, they can demand to study or focus more on ABC. (Elif, I1)

Thus, Elif believes that students can also make use of assessments and improve their language skills. In the same vein, Esra thinks that students see the path that they need to follow with the help of assessment techniques. Also, she highlights that assessment is a way to produce the language they have been learning.

It is really important, and it is a need. Both students and teachers need assessment because students also need to see what they can or cannot do. Therefore, it also shows the path to teachers and students, which is very important. (Esra, I1)

Also, assessment helps students produce the language they learn, so it is a great opportunity for them, I think. Generally, they are scared of testing themselves, but it is really helpful because they need to see what they can do. So, this is a great production way. (Esra, I1)

In addition, Esra guides her students and holds them accountable for their own learning by providing them with feedback and assigning them extra work.

I agree with my colleague Elif, and sometimes I advise students and say, “Okay, you need to focus on this topic. Study more on this topic, please, after the lesson because you have a lot of mistakes.” After a writing task, I give such feedback. [...] (Esra, I1)

Songül also supports the idea that students can use assessment as a helpful tool even though it might negatively affect some of the students’ psychology.

I think [students] are fixated on the idea that they need to pass or fail, but at the same time, I see some students who are more conscious about education. They try to follow up the idea that “I had 10 in the first exam, I have 15 in this one, so I am showing progress.” [...] (Songül, I3)

Finally, Ekin believes that assessment practices, such as portfolio assignments, are highly positively influential on student learning. It should also be noted that portfolio assignments include self-reflections in her institution.

[...] We are doing writing and speaking tasks in our school as a part of the assessment. I mean, portfolio. It is a part of the assessment as well and has a huge impact on students’ learning. [...] (Ekin, I2)

To summarise, this section has provided instructors’ opinions on how assessment can be a positive component for students in their learning process. Similarly, Schellekens et al. (2021)

indicated that students were required to be self-regulated by the comprehension of the purpose of assessment and evaluation of their own work. Therefore, the findings of this study demonstrate that students are expected to study the topics suggested by their teachers, reflect on their exam results, and decide on their future actions to improve their own learning. Furthermore, it is important to make students aware of the fact that they can benefit from assessment, rather than being fixated on their exam scores all the time.

4.1.3.3 Proficiency Exam as a High-Stakes Exam

As previously stated, student accountability conception is about gathering information about students all the time and comparing those results with the help of a rubric, students' previous performances, and other students, which result in significant consequences for them, such as graduation (Brown, 2008). The process of making students take exams and defining their level based on a rubric or other students' performances can be related to the proficiency exams applied in preparatory schools in the context of this study. Proficiency exams, which take place either at the beginning or at the end of each school year, are crucial for students since they affect their lives. Hence, the opinions of instructors on the proficiency exam as well as some information on how they use proficiency exam results in the framework of student accountability will be presented in this section.

Firstly, Melda, who has been teaching in the same institution for more than ten years, points to an institutional practice. She states that students' performances are used to make decisions about them, such as changing their classes when necessary.

In my institution, we have a slightly different approach. If students cannot reach the goals of a level, we generally use those assessment results and change their level. We kind of demote them to a lower level, and we also use students' successful results and give them another chance to move up one more level [...], so we use assessment results like this in addition to what has been said. (Melda, I2)

Ebru also finds this method of changing students' levels based on their performance effective in the context of preparatory schools.

I wish we had something similar to what Melda said. Sometimes, after the placement test, students are placed in a classroom but do not feel comfortable in that classroom either. It can be because of their higher or lower level. Especially this term, [...] in my beginner-level classroom, there were some students who were pre-intermediate and some students who could not communicate in English, and it was a very big problem for me. [...] (Ebru, I2)

On the other hand, Songül and Hatice do not think that their institutions' proficiency exam procedures are helpful when deciding students' English levels. Songül's remarks are parallel to those mentioned by Ebru.

[...] If they actually prepared exams based on A1, A2, B1, B2 or whatever, they would not have to mix and match all the levels and make a D level, C level, B level, or we would not have students who cannot even write a sentence describing themselves and a student who can talk about environmental crisis. So, we would not have that kind of students in one classroom, I think. (Songül, I3)

Hatice's statement highlights the standardisation problems that can be faced in the proficiency exam. Hatice's idea of standardisation problems is also agreed upon by Elif, although she thinks that the proficiency exam is a good indicator of students' levels in her institution.

[...] I guess the results of the proficiency exam do not reflect the students' real level because, as we know, there are some parts that assess students' level [in this exam]: writing, speaking, and mechanical parts. The mechanic test might be objective, but the written and the spoken assessments are more likely to be instructors' way of assessing. At that point, I guess there must be a standard for teachers' perspective because I might think of giving a score of 15 for a writing paper, but the same paper can be assessed as 25 [by another instructor.] That is a huge gap between the two instructors, so at that point, there must be a standard, and I do not believe that there is a perfect standardisation for the assessment of writing or spoken exams. [...] (Hatice, I3)

We test all the skills in the proficiency exam, and I believe it gives a high percentage of correct results, but of course, when there is a writing and speaking assessment, there are the assessors, of course, and there is not just one person assessing all the students, so standardisation is always a problem, I think. No matter how we study the subject and have those meetings to grade those papers or the speech, there could be a weakness in that sense. I mean, if somebody comes and examines our results, I can say that this could be our weakness. I think it is a general issue, not for our institution. The moment writing or speaking is involved and assessed by someone, of course, there are discrepancies. (Elif, I1)

Overall, by considering instructors' opinions and criticisms, proficiency exam practices can be improved, making student accountability more meaningful. Furthermore, according to Smith (1991), teachers believe the negative emotional influence of high-stakes exams on students by indicating anxiety or worry. Besides, high-stakes exams are demotivating, especially for lower achievers (as cited in Brown, 2008). Similar opinions were evident in the focus group interview sessions in this study and can be mainly framed around the proficiency exam and student accountability conception.

4.1.3.4 High-Stakes Assessment and Student Emotions

It is indicated by Brown (2008) that focusing on students' emotions and well-being as well as creating a positive classroom environment may be factors leading teachers to consider assessment irrelevant. Thus, the opinions presented below can be interpreted in the light of student accountability and irrelevance conceptions together.

First, Buket shares her opinions on students' stress concerning assessment. Moreover, she confesses feeling some stress due to assessment as a teacher. Below are two excerpts from different parts of her speech. In the first one, she explains the reasons for the negative feelings of students while concluding her speech by saying that teachers should change their

perspectives, which will pave the way for a positive understanding of assessment. In the second excerpt, she indicates that students would be happier in a world without assessment; however, she also uses the subject “we”, most probably to indicate teachers themselves.

The first thing that comes to my mind [about assessment] is actually negative when I think of the student perspective because when you are being assessed, there is a kind of burden on you, and you have to be good at it, so the first thing that comes to my mind is negative... Maybe we should change our perspective first as instructors and make students feel comfortable about being assessed. Then they will also feel more relaxed, I think. (Buket, I3)

[...] I think this is not only the exams, but everything about assessment that gives us this idea [of if students have learned], so in-class assessments, etc. give us the sense of achievement in the end, and this part is important for me. But actually, if there were no exams and things like that, we would be more relaxed. I think it would be even... This might be bold, but we would be more... The students would feel happier to learn in the end, maybe. (Buket, I3)

In addition to Buket, Esra and Songül underline the negative emotions students have towards assessment. Because of these negative emotions, such as stress and anxiety, Songül points out that some assessment procedures should be changed.

Generally, they are scared of testing themselves, but it is really helpful because they need to see what they can do. So, this is a great production way. (Esra, I1)

[...] I think we should make [assessment] a bit more individualistic and softer than what we are doing right now because it makes me sad to see students getting stressed over some useless preposition items in a long phrase that they have to memorise [...] (Songül, I3)

Songül, elsewhere in her speech, states how students are focused on the idea of the final exam they will take at the end of the term/year and get nervous about it, although this might not be the case for all the students.

Well, I think students are usually stressed about assessments, and the only focus they have until the end of term is this big final exam, midterm exam, and the presentation they have. I think [students] are fixated on the idea that they need to pass or fail, but at the same time, I see some students who are more conscious about education. They try to follow up the idea that “I had 10 in the first exam, I have 15 in this one, so I am showing progress.” [...] (Songül, I3)

Leyla thinks that while assessment can be perceived as positive for some students, some consider it negative because they are afraid of failing or getting low grades. This statement is discussed in the framework of student accountability conception since it is related to categorising students based on their grades and choosing or certifying them.

[...] I believe that students, most of the time, look at testing as a way to be positively reinforced or negatively reinforced, so they are either very, very afraid of getting a bad grade and failing, or they are very ambitious and want to get good grades and have this positive end. Everything that they do along the way to pass the proficiency exam at the end is in service of that. (Leyla, I1)

As given in the aforementioned sections, Ebru included her thoughts on how different assessment methods can positively or negatively affect students’ feelings by touching upon her past experiences in a different institution (See page 75). It should be noted that she also felt confident working there as a teacher, indicating high-stakes exams’ negative effects on students and teachers.

[...] I was teaching at a language school, and it was just a portfolio assessment. We did not have any written exams, and I felt confident about it because the students were there to improve their English. They had some realistic expectations and when we did the portfolio assessment, they were willing to participate, and they did not have a score anxiety, let me put it in that way. They wanted to learn English. [...] (Ebru, I2)

Clearly, scores and exams are sources of anxiety for learners. Similarly, Hatice believes that some types of assessment can cause stress for students by making them nervous. On the other hand, Hatice also thinks that grading can be a motivating factor for students. Thus, it can be interpreted that assessment is necessary and can be a driving force by putting some students under pressure and making them stressed.

[...] I believe that the only way of assessing students is not the written way or making them scared of assessment. [...] (Hatice, I3)

[...] If they are aware of the situation that they will be assessed at the end of a term or a year, they will push themselves even if they do not want to. So, I guess, to some extent, assessment can encourage learners to check what they have learned in a class. (Hatice, I3)

In short, most of the instructors participating in this study are aware of the negative emotions that exams evoke in students, although this is not the case for a minority of them. Therefore, it can be concluded that the focus on high-stakes exams, as well as a high level of student accountability conception, can result in negative feelings attached to assessment, as repeatedly expressed by the participants. In this vein, Harris and Brown (2009) pointed out that some teachers in the New Zealand context did not consider student accountability conception valid in assessment since competition among students could demotivate some students, especially lower achievers. Moreover, the opinions of the participants are aligned with the statements of Flores et al. (2015) who found out that written exams did not seek ways to improve the learning process by informing teachers, but they were rather about punishing mistakes.

Overall, holding students accountable for their own learning is preferred by the instructors participating in this study. They try to achieve that by using self-assessment and peer assessment strategies or by guiding their students to study individually, as exemplified above. However, when combined with the high-stakes exam context, students, and even teachers, may feel under pressure from time to time. If these negative emotions and pressure are too much, it can even lead to perceiving assessment as irrelevant.

4.1.4 School Accountability Conception

Finally, descriptive statistics show that the last conception that the participants hold in this study is the school accountability conception ($M=11.55$). Even though this conception has the last place, the amount of agreement is not that low. This can be interpreted as the fact that many

instructors do not totally consider assessment as a way to demonstrate their success as schools or teachers, but instead, they are more concerned with the improvement and student accountability as well as the problems they mention regarding the irrelevance conception.

As clarified in the literature review section, school accountability is a way for schools and teachers to demonstrate that they have done a successful job (Brown, 2008). Hence, the ones believing in this conception may have a direct link by calling institutions and teachers successful if students are successful in exams. In this study, the participants were requested to share their opinions on a specific question regarding school accountability conception. In other words, based on the items concerning school accountability on the TCoA-IIIA, such as “*Assessment is an accurate indicator of a school’s quality.*”, participants were asked a similar question to reflect on. Primarily, instructors cannot see a direct link between students’ success and institutions’ success.

4.1.4.1 No Direct Link Between Institutional/Teacher Success and Student Success

When the participants were asked if they agreed with the positive relationship between the success of students and institutions or teachers, they implied that they partially agreed with this opinion. The common point that all the instructors suggest is that the influences of some external factors cannot be omitted when it comes to labelling schools as “good” or “bad.” Almost all of them discussed that many other factors might affect student success rather than the success of institutions or teachers. Therefore, no direct link was found between these two elements in this study. Firstly, Elif emphasises the external factors, such as the problems institutions have regarding exam questions, which makes it difficult to have a direct parallelism.

In an ideal world, theoretically, or if somebody writes a book about it, assessment results could reflect the success of teachers and the success of the institutions, but maybe the testing office asked a really easy question in that exam; or the opposite, most of the students failed because the testing person was really angry that month, you know, it happens. The teachers may have other problems in the institution, and they could not focus on teaching for a while. That does not show their potential but shows other problems, so we cannot make a direct link from here to there. But of course, assessment shows something about teaching if we do not count any other external issues. (Elif, I1)

Ebru, on the other hand, offers a different perspective and claims that teachers’ positive attitudes can affect students in a positive way, which may lead to success. Nevertheless, it should also be noted that sometimes, even when teachers try very hard to improve their students’ learning, students do not respond in the expected way, as Ebru states. This makes it difficult to point a direct link again.

They can be related to each other, but sometimes, it does not matter how successful or hardworking a teacher is. Sometimes, teachers are more hardworking than students; I had classes like that. They did not care about their results. I reminded them of the upcoming exam and prepared so many tests for them to practise more, but they did not take them seriously. But of course, the success of the teacher affects students. I mean, teachers are role models for the students, right? Therefore, when the teacher is a very successful and hardworking one or if s/he is very willing to teach English or [...] try to teach them strategies, "You can do this by looking at this tip." or "We are using this strategy, this is the short way for this type of questions." so that students can understand better, I mean, they are more aware of the strategies, right? [In this way, they can affect their students.] That is why they affect each other. I feel like that. (Ebru, I2)

Additionally, Ekin feels that institutions and teachers are responsible for certain aspects concerning assessment, even though she believes that student assessment results are only sometimes a direct indicator of institutional success, even if those responsibilities have been carried out well.

There are some other things as well. Students' assessment results do not only depend on a school's success. They are not 100% connected, but to a certain extent, they are connected. I mean, can you give your objectives, or is your curriculum working? Do you have realistic expectations and high-quality teachers? These are the things that students' results may show, but the results may not show it either. Of course, if there is a chronic decrease in the results or all these results are low, there is a problem with this situation. But we cannot say it is the institution's problem or its unsuccessful practices if some students' results are low. (Ekin, I2)

The other participants concur with their colleagues' opinions. They believe that school accountability depends on many factors. However, while clarifying their points, they pay attention to different issues. Specifically, Leyla states that success is a vague term that can change from one institution to another. Therefore, she believes that institutions cannot be called "successful" or "not successful" if students show a certain level of success. Besides, she thinks that sometimes students fail despite all the efforts of teachers.

I think you can roughly call a teacher successful if a high percentage of students get above a certain grade, but not an institution because institutional practices change quite a lot. [Also,] "successful" in one institution's terms and one institution's assessment criteria might be very unsuccessful in another, but you can say that this teacher has done a good job or a bad job. But I think you have to take care of the fact that certain groups sometimes have certain behavioural issues. They have certain linguistic issues, and sometimes they just do not get it, and it is not the teacher's fault. Again, there are classroom dynamics in place here, so you can carefully assess the teacher but not the institution, I think. (Leyla, I1)

Songül explains the same reasons as Elif. If institutions and teachers perform their responsibilities regularly and carefully by considering curriculum and preparing well-designed exams, then students' failure might be about their performance, which is not associated with teacher or institutional practices.

There can be a relationship because if the institution prepares an impossible exam, it is the institution's fault. If the teacher is being too harsh or too dismissive of the curriculum while grading papers, the teacher is going to be accountable for that. But other than that, if everything is going normal, if everyone is doing average on tests, then students should look at themselves and see if there is a problem with their own understanding of an assessment and education in general, I think. (Songül, I3)

Finally, Hatice approaches the topic from a different perspective. She believes that students' own studies or background knowledge in English affect their performance and that institutions or teachers are not the only ones accountable for their success.

I guess it might not only be the institutions' or the teacher's success [if students perform well or not.] For example, if a student takes a proficiency exam before taking their courses, and this student becomes very successful and has a B2 level, it is not because of the institution's success. So, [...] if I am a new student, it is only my success or the reason is my background in English learning, not that institutional effect. (Hatice, I3)

In short, none of the instructors indicate that assessment is a way to label schools or teachers as quality ones or not based on students' performance. This is because they believe that many environmental factors should be considered.

4.1.4.2 Solutions to Diminish the Feeling of Incompetence

Teachers are one of the significant pillars of the assessment process, as implied by Brown (2008). Therefore, their accountability for assessment is meaningful. Concerning this, another fundamental element in this conception is related to improving schools and teachers. In other words, high-stakes assessments lead to improvements in assessment and teacher quality as they are held responsible (Brown, 2008). Concerning this, Lutovac and Flores (2022) discovered that PSTs also thought that the role that teachers played in students' success and failure was crucial by holding teachers responsible for students' failure. Therefore, they believed that teachers should revise their teaching and assessment methods for improvement purposes. In the Turkish context, teachers are also held accountable for students' success or failure, which was reflected in the participants' statements during the interviews. To this end, many participants highlighted the importance of cooperation between teachers as well as in-service training since most participants state that they feel they need to be more competent in their knowledge about assessment. While some of them mention that they need to improve themselves regarding in-class assessment, some participants need more training in preparing quality exam papers. Their statements can be interpreted as holding themselves accountable as teachers and looking for ways to improve their competencies. To begin with, Elif believes that cooperation among teachers and the institution's testing office will help her feel more confident.

Well, collaboration, I think because we are not alone, as you can see [in the interview.] Everybody has the same problem. I do not want to say, "Okay, let us go and read articles." because I do not find it very practical. However, I think the instructors should come together and do research or share experiences in their institutions. Maybe they can read and research together, especially with the testing and assessment team because we are doing some kind of assessment in the class, and that is fine. However, at the end of the day, we prepare our students for a specific type of assessment in each institution. [...] It is not the solution, maybe. After one year, no one will learn everything, but it will ease the issue in time. We can feel more and more confident after years. (Elif, I1)

Furthermore, Leyla is mostly in alignment with Elif. She thinks that classroom dynamics are quite different from each other, and instructors should consider these differences regarding in-class assessment. For this aim, communication and cooperation between teachers are also said to be crucial points, as also argued by Yan and Fan (2021).

I agree with what Elif said, and I think every institution and classroom within that institution has a certain dynamic. At one point, everyone needs to touch base and know that this certain group is not doing well with in-class feedback, and we need to spare some time to give them feedback individually, for example. I think communication between teachers who are teaching the same class and those who are teaching at the same level in the same term will help because most of the time, teachers are very isolated, even from their partners because we just do not have the same schedules. [...] These are very necessary because I have had times where when I spoke to students in class and said, "No, you are still doing it wrong. You need to say, I am going." for instance. The student and the class responded very well, and it went great. But in another classroom, as soon as I made that mistake of fixing the error, the students shut down and did not speak for the rest of the class, so those dynamics are very important. All teachers need to come together and talk about their observations. (Leyla, I1)

Melda has been working as an instructor for a long time, just like Elif. In her case, she assigns so much importance to the in-service training as she thinks assessment is a field that requires constant research and development. It is concluded from her first excerpt that she might need more training if she changes her duties regarding assessment in her institution. Moreover, in the second excerpt, she delves into the necessity of improving assessment knowledge and practices continuously, which can be acquired through training or collaboration.

[...] If I had such a role in my institution, I would probably need training and maybe some self-study. (Melda, I2)

As far as I think, one course is not enough. But assessment is something just like teaching, maybe; you should feed constantly and get training. You should work with others, observe others, and work with people from those committees or groups. You need to feed it just like teaching. In that sense, I believe teaching and assessment are quite similar. Sometimes, things change. Two or three years ago, we had COVID, and all of a sudden, we did online assessments. We did not know anything about online assessment. Some universities did it, but normal people did not. And we had to learn it. After that, our views and perspectives changed, so it is something that constantly develops and improves. (Melda, I2)

Ekin has been working in the testing office at her institution for a few years. Although she is not an ELT graduate, she has improved herself with the help of in-service training. Thus, the positive effect of in-service training cannot be denied in her case as well. Furthermore, her statements indicate that teachers should improve themselves continuously, similar to Melda.

I am not an ELT graduate, so I had not taken testing courses during my bachelor's years, but when I started to do testing, I started to get some in-service training. So, I learned the job both there and by doing it, and I am still learning and still have a lot to learn. (Ekin, I2)

Lastly, during the third interview session, Hatice stated how difficult it was for her to apply assessment techniques for the first time as a teacher. She believes that if she had been provided with some training in cooperation with the teachers who had been more knowledgeable in assessment, she would have felt much better about her competencies.

[...] I remember my first year teaching at a university, so I had not had any experience in assessing my students' level in any way, and it made me nervous and stressed. Maybe at those times, some specific or short-term courses could have been given to in-service teachers, so even if the teachers had not had any courses at their undergraduate level or postgraduate level, they might have been experienced thanks to these courses given by universities because I strongly believe that if I was given a course, I could have been better. Of course, this course should have been given by teachers who were experienced in testing and assessment, so I could observe what to do. [...] Thus, observing experienced teachers and how they assess their students [would have been better.] (Hatice, I3)

All in all, teachers held themselves accountable for student learning and were willing to improve their competencies, which is related to both improvement and school accountability conceptions.

4.1.4.3 Washback Effect

Berry (2008) addresses the negative backwash effect concerning high-stakes exams by underlying their negative and narrowing effect on student learning. Additionally, while delving into the details of school accountability conception, Brown (2008) points out that high-stakes exams and their emphasis can lead teachers to focus mainly on exams, which he calls “teaching to the test.” It can be understood that if teachers and institutions are responsible for students' success and use exam results to show their success publicly, exam scores can be their main concern, which is also acknowledged by Yan, Zhang, and Fan (2018). In this study, the participants mentioned the washback effect; however, they did not imply a personal change in their focus on exam-related content. They mainly criticised their students who were influenced by the washback effect. Below are some instructors' opinions about the effect of exams on teaching or learning. Firstly, Ebru mentions how assessment can negatively affect student learning.

It may have some positive effects and negative effects. Well, when I say negative effect, for example, students do not have to pass a speaking test to pass the preparatory school in my institution. Therefore, they want to focus on grammar, vocabulary, and reading more than doing some speaking exercises, [but] it is not learning English. I mean, learning English, they lack the skill of speaking most of the time just because of that. (Ebru, I2)

Moreover, she goes on to state that students may be more prone to learning a topic if they know that they will see it in an exam, which she calls the positive effect.

[...] It has a positive effect, especially when we are teaching the grammar part or the vocabulary part. Students are more focused because they know that they are going to use this information for a big purpose, which is passing the preparatory school exam. Therefore, they pay more attention when it is something related to exams. (Ebru, I2)

During the same interview session, Melda mentions the same points in the following words. She observes that students tend to concentrate on the topics that they will encounter in their exams.

[...] If something is not in the test, students generally do not study and pay attention to it. Also, generally, this is what I have seen; assessment given by other peers does not generally affect their learning. Maybe they do not trust their peers. Sometimes, they do not trust themselves, and they may say, "How should I evaluate my own learning?" From that perspective, they pay more attention to the feedback or the score coming from the professor or the centralised test. (Melda, I2)

Ekin, as an instructor working in the testing office, is well aware of assessment and how it should be done. While explaining her thoughts on assessment, she also touches upon the washback effect.

[...] The important thing here is that: do your assessment methods align with your curriculum or do you try to make your curriculum align with your testing method? I mean, if it is the first one; if your assessment methods are based on your current curriculum, it means you are on the right track to help your student with the results of your assessment. (Ekin, I2)

Besides, Esra's sentences might be in line with the washback effect as well since she mentions that assessment practices lead her to consider her teaching policy accordingly.

Assessment helps me to see how I can teach also, so now I am thinking about teaching when I think about asking something, for example. How can I teach it and how can I ask it? So, there is a link between these two things. (Esra, I1)

Lastly, Brown (2008) points to the negative consequences of holding school accountability conception because of the stress high-stakes exams evoke in teachers and students. The words Buket utters are aligned with this point. These excerpts have already been included under the student accountability conception and can also be discussed in the framework of the irrelevance conception.

The first thing that comes to my mind [about assessment] is actually negative when I think of the student perspective because when you are being assessed, there is a kind of burden on you, and you have to be good at it, so the first thing that comes to my mind is negative... Maybe we should change our perspective first as instructors and make students feel comfortable about being assessed. Then they will also feel more relaxed, I think. (Buket, I3)

[...] I think this is not only the exams, but everything about assessment that gives us this idea [of if students have learned], so in-class assessments, etc. give us the sense of achievement in the end, and this part is important for me. But actually, if there were no exams and things like that, we would be more relaxed. I think it would be even... This might be bold, but we would be more... The students would feel happier to learn in the end, maybe. (Buket, I3)

In short, the influence of assessment on the learning environment is undeniable and a topic of concern mentioned by the participants of this study. As Azis (2012) pointed out, when teachers held school accountability conception, they tended to turn their attention to demonstrating the quality of instruction provided by them or by the institution, along with improving this quality. Thus, it can be concluded that when teachers are mainly concerned about demonstrating their success, they might narrow their focus by sticking to the exams, which is called the washback effect, although the participants referred to the washback effect on the part of students instead of their own perspectives.

Furthermore, the findings concerning school accountability conception demonstrated that no direct connection between the success of students and teachers/institutions was found by the

participants and school accountability had the last place. This finding is different from the findings of Yüce (2015), who underlined that improvement and school accountability conceptions had the highest mean scores in her study. As Yüce (2015) investigated the assessment conceptions of PSTs, one possible interpretation can be that improvement conception does not change over time through experiences as in-service teachers. However, this is not the case for school accountability conception, as it holds the last place in this study.

Overall, the findings of the study highlight the improvement and irrelevance conceptions, while student accountability and school accountability conceptions were also valid in explaining the conceptions of the participants, even though the mean scores of these two conceptions are relatively low.



4.2 RQ1.1: How do these conceptions vary according to instructors' experience years?

To answer research question 1.1, comparison analyses were conducted by using the scores of the four subscales. Initially, the Shapiro-Wilk test was run to test normality. In other words, Shapiro-Wilk test is helpful in understanding if the distribution of a data set is normally distributed. It starts with the null hypothesis that the data has a normal distribution. If the p-value is low, it means that the hypothesis is rejected, and the data does not follow a normal distribution. Thus, a large p-value is needed for a normal distribution (Malato, 2023). As the normality assumption for the score variables is fulfilled for each group regarding participants' years of experience ($p\text{-value} > 0.05$), ANOVA was conducted, and the differences between groups were analysed in a parametric way.

Table 11

Tests of Normality (Shapiro-Wilk Test)

Tests of Normality	How long have you been teaching?	Shapiro-Wilk	
		Statistic	Sig.
Student Acc. Score	Less than five years	0.946	0.287
	Between 5-10 years	0.937	0.347
	More than ten years	0.965	0.066
School Acc. Score	Less than five years	0.959	0.493
	Between 5-10 years	0.942	0.407
	More than ten years	0.968	0.086
Irrelevance Score	Less than five years	0.972	0.771
	Between 5-10 years	0.940	0.388
	More than ten years	0.977	0.257
Improvement Score	Less than five years	0.982	0.950
	Between 5-10 years	0.948	0.490
	More than ten years	0.968	0.087

After the normality test, Analysis of Variance (or One-Way Analysis of Variance) was conducted since it is used in studies with a normal distribution to comprehend the differences between three or more groups (Otrar, n.d.). ANOVA is one of the most commonly used statistical practices, as indicated by Dörnyei (2007). The results of ANOVA are presented below.

Table 12

ANOVA Results

Subscales	Experience Years	N	Mean	Std. Deviation	F	Sig.
Student Acc. Score	Less than five years	21	13.33	2.497	2.585	0.081
	Between 5-10 years	15	13.47	2.850		
	More than ten years	65	12.23	2.350		
School Acc. Score	Less than five years	21	12.05	2.958	0.557	0.575
	Between 5-10 years	15	11.80	2.242		
	More than ten years	65	11.34	2.933		
Irrelevance Score	Less than five years	21	32.90	5.770	2.626	0.077
	Between 5-10 years	15	28.00	7.010		
	More than ten years	65	29.72	6.997		
Improvement Score	Less than five years	21	50.57	7.620	2.683	0.073
	Between 5-10 years	15	55.60	7.414		
	More than ten years	65	49.66	9.626		

The findings suggest that as the p-values for each test are $p > 0.05$, the difference between the average scores of the groups is not significant. Looking at the assessment conceptions of instructors with less than five years of experience, they mainly have the *improvement conception* ($M=50.57$). This is followed by *irrelevance conception* ($M=32.90$). *Student accountability conception* ($M=13.33$) is the third one preferred by instructors with less than five years of experience. Although *school accountability conception* has the minimum weight ($M=12.05$), the difference between student accountability and school accountability conceptions is slight.

Secondly, when the experience years of the instructors are between 5-10, *improvement conception* also has the highest value ($M=55.60$), which is slightly higher than those with less than five years of experience. Secondly, instructors in this group also have the *irrelevance conception* following the improvement conception ($M=28.00$). The reliance on irrelevance conception is slightly less than the first group, even though it is not a big difference. School accountability and student accountability follow the same pattern within the group of instructors with between 5-10 years of experience. While *student accountability conception* is in third place ($M=13.33$), *school accountability* follows it with the lowest score ($M=11.80$).

Surprisingly, based on the analysis of the last category, which includes instructors with more than ten years of experience, *improvement conception* has the highest score ($M=49.66$) despite

being the lowest one among these three categories of experience years. Again, *irrelevance conception* is in the second place ($M=29.72$). Similarly, *school accountability* ($M=11.34$) succeeds *student accountability conception* ($M=12.33$).

In other words, ANOVA results demonstrate that the differences in the experience years of instructors do not affect their assessment conceptions significantly. This finding is aligned with the findings of Brown (2002). He also found that participants' years of experience did not lead to considerable differences in assessment conceptions (as cited in Vardar, 2010). Moreover, the findings of Ma and Bui (2021) followed a similar direction. They highlighted that teachers' years of experience did not have a significant effect on their assessment conceptions in the Chinese context. Regarding the Turkish context, similar results can be seen, as demonstrated by some studies in the literature (Önalan & Karagül, 2018; Zaimoğlu, 2013). Therefore, it is possible to reach a conclusion that the years of experience do not affect the assessment conceptions of teachers by also taking the findings of this study into consideration.

More precisely, improvement conception has the highest value for the instructors in all the categories in the present study, while it has the highest mean score with teachers whose experience years are between 5-10 years. ($M=55.60$) Instructors with more than ten years of experience have the lowest improvement agreement, even though the distinction is not noteworthy ($M=49.66$). Therefore, it was found that all the instructors considered assessment constructive, leading to positive changes in students' learning. The slightly lower agreement level of the most experienced instructors may be related to their being more accustomed to traditional ways of assessment. Additionally, irrelevance conception has the highest level of consensus among instructors with less than five years of experience ($M=32.90$), and this is similar to Vardar's findings (2010). She pointed out that it might be related to the lower quality of courses concerning assessment. When it comes to student accountability and school accountability conceptions, the discrepancy between these two conceptions is also not noticeable.

Nevertheless, to gain a profound understanding, the participants in the focus group interview sessions were also requested to share if their perceptions of assessment have changed through gaining more experiences as in-service teachers. Many instructors indicated that they could observe a change in their understanding of assessment as time passed; however, they responded to this question by approaching it from different perspectives. For instance, Melda stated that her teaching and assessment philosophy has changed a lot with the help of her experiences.

[...] Your views about teaching and assessment change as you get more experienced. I remember my first year as a teacher. My teaching philosophy was totally different. My assessment idea was totally different. So, as you get experienced your views change a lot. [...] (Melda, I2)

Regarding this, Melda and Ekin also pointed out that assessment was dynamic and needed to change through the needs of the time, which required continuous effort from the teachers' side to keep up with these changes. Thus, they believed that their perception of assessment has changed and will continue to change inevitably.

As far as I think, one course is not enough. But assessment is something just like teaching, maybe; you should feed constantly and get training. You should work with others, observe others, and work with people from those committees or groups. You need to feed it just like teaching. In that sense, I believe teaching and assessment are quite similar. Sometimes, things change. Two or three years ago, we had COVID, and all of a sudden, we did online assessments. We did not know anything about online assessment. Some universities did it, but normal people did not. And we had to learn it. After that, our views and perspectives changed, so it is something that constantly develops and improves. (Melda, I2)

[...] The trainings that I got shaped my understanding and my philosophy, especially about the alignment of objectives and the exams themselves. However, the thing is that the philosophy of the world about assessment is changing, so it will constantly change. The trends will change, the ways that we ask questions will change, the way we give feedback will change, so it will never remain the same and the effects that we get from the trainings or the courses that we got will not stay on us. (Ekin, I2)

Leyla, on the other hand, mentioned that she began to detect some patterns in her students' mistakes, leading her to reach better conclusions about her students and implement more meaningful strategies to improve her students' language learning process.

Perhaps, because of the lack of instruction about these things in university, I thought that testing was more about finding mistakes. [...] However, it was towards the middle of my very short teaching career that I realised certain mistakes reflected certain patterns and we could make deductions and think about better ways to perhaps instruct [the necessary points.] In my first teaching experience, I could detect the mistakes, but I could not really understand that there was a pattern most of the time- now I do, I think. (Leyla, I1)

Elif, who took part in the same interview session, stated that she gained insight into the fact that assessment was not the direct result or result of teaching and learning through her experiences and observations of students. Also, while describing what assessment meant to her, she indicated that assessment demonstrated the future actions that she needed to apply: revision or moving on. This may imply that she prefers making changes in her teaching, so assessment is not the endpoint, but it includes the revision process for improvement purposes. A similar point of view was included in Schellekens et al. (2021) while mentioning AfL. They implied that assessment and feedback were not the end point but should be the point where the learning process starts.

[...] Through the years, I learned that assessment is not the direct result or the end result of teaching or learning and of course, I try to assess the students continuously, not just with pen and paper or portfolios or something formal, but I also observe them and ask them questions. (Elif, I1)

Overall, the aforementioned statements of participants in the focus group interview sessions have parallelism with the statements of Bonner (2016), indicating that assessment practices and

conceptions could gradually alter via professional development (as cited in Comforto, 2022). However, it is essential to emphasise that even though some opinions and concerns of instructors have changed over time, they have not experienced a dramatic change, such as a conception change from improvement to irrelevance, for instance, after ten years of teaching experience. Therefore, it can be deduced from the focus group interviews that instructors' opinions on assessment have changed to some extent, although the years of experience do not make drastic differences in their conceptions, as also presented by the quantitative data results.



4.3 RQ1.2: How do the assessment conceptions of the instructors differ based on their educational background?

To answer research question 1.2, Shapiro-Wilk test was run first to see whether the data concerning departments of the participants was normally distributed. The table related to the results of Shapiro-Wilk test is presented below.

Table 13

Tests of Normality (Shapiro-Wilk Test)

Tests of Normality	Which department did you graduate from?	Shapiro-Wilk	
		Statistic	Sig.
Student Acc. Score	ELT	0.963	0.041
	ELL	0.962	0.496
	Other	0.953	0.703
School Acc. Score	ELT	0.950	0.008
	ELL	0.961	0.479
	Other	0.912	0.295
Irrelevance Score	ELT	0.979	0.320
	ELL	0.949	0.285
	Other	0.948	0.644
Improvement Score	ELT	0.957	0.020
	ELL	0.977	0.841
	Other	0.954	0.713

Since the data normality assumption for the score variables is not fulfilled for all groups ($p\text{-value} < 0.05$), the analysis was completed using the Kruskal-Wallis H test, which is a non-parametric test that can be utilised when the data is not normally distributed. Also, the Kruskal-Wallis test is described as an alternative to ANOVA (Dörnyei, 2007). In other words, since the number of observations in the subscales is low in the present study, the comparisons of differences between subscales were tested with the non-parametric Kruskal-Wallis H test.

Table 14

Kruskal-Wallis H Test Results

Subscales	Department	N	Mean Rank	Kruskal-Wallis H	Sig.
Student Acc. Score	ELT	68	53.57	1.686	0.431
	ELL	23	44.93		
	Other	10	47.45		
School Acc. Score	ELT	66	48.28	0.712	0.701
	ELL	23	57.55		
	Other	10	51.63		
Irrelevance Score	ELT	68	35.90	3.292	0.193
	ELL	23	53.65		
	Other	10	42.46		
Improvement Score	ELT	67	53.57	2.547	0.280
	ELL	23	44.93		
	Other	9	47.45		

Based on the findings presented above, the p-values for each test are $p > 0.05$, which can be interpreted as the difference between the scores of the groups not being significant. First, instructors who graduated from the ELT Department mainly chose items about *student accountability* and *improvement conceptions* ($M=53.57$). While the *school accountability* score is the third conception they prefer ($M=48.28$), the *irrelevance conception* has the lowest mean score ($M=35.90$).

On the other hand, the preferences of instructors who graduated from the ELL Department differ from the previous answers, although the numbers change insignificantly. For instance, the first conception of the instructors who graduated from the ELL Department is *school accountability* ($M=57.55$), followed by *irrelevance* conception ($M=53.65$). After these conceptions, *improvement*, and *student accountability conceptions* have the same mean score ($M=44.93$).

The last category, Other, consisted of instructors who graduated from American Culture & Literature, English Translation & Interpretation, and Linguistics. They were placed under the same category because there were few participants in each group to reach meaningful results. Looking at the results, *school accountability* has the highest mean score ($M= 51.63$), the same as the instructors who graduated from the ELL Department. However, *student accountability* and *improvement conceptions* share second place with the same mean score ($M=47.45$). The

mean score of *irrelevance conception* (M=42.46) has the lowest mean score in this category, just as it did for instructors who graduated from the ELT Department.

To put it differently, the Kruskal-Wallis H test results demonstrate that even though the differences between subscales do not differ remarkably, improvement (M=53.57) and student accountability conceptions (M=53.57) have the highest mean values in teachers who graduated from the ELT Department, similar to the findings of Zaimoğlu (2013). On the other hand, irrelevance conception has the smallest agreement by the graduates of the ELT Department (M=35.90), while it is more agreed upon by the instructors who graduated from the ELL Department (M=53.65), together with the graduates of the Other category (M=42.46). This may imply that the graduates of the ELT Department are slightly more aware of the importance of assessment, which might be related to the pre-service education they were offered. However, it is also worth noting that the difference is not remarkable. Thus, whether the participants had pre-service courses related to assessment or not does not make meaningful changes, which can be discussed with the comments of the participants taking place in the focus group interview sessions.

During the interviews, most participants indicated that they did not feel competent concerning assessment, similar to the findings of the previous studies. Pastore (2023a) pointed out that the participants in his/her study also mentioned their assessment illiteracy. Similarly, Stiggins (2004) indicated that only a very small percentage of teachers felt ready for classroom assessment as they did not learn how to do it. Furthermore, while reflecting on their pre-service education, many of the participants in this study mentioned that they either did not find those courses enough to equip them with the knowledge they needed or that these courses gave them an incorrect perception of assessment and teachers as assessors. Some of the reasons for the insufficient education of PSTs were mentioned by Vardar (2010). It was stated that these reasons were the theoretical information provided to teachers rather than the practical knowledge, which was covered in crowded classes. Supporting the findings of this study, Güngör and Güngör (2024) indicated the lack of assessment courses in teacher-education programmes. Although the instructors in this study did not say anything about the number of assessment courses, the lack of practical knowledge and having been bombarded with theoretical knowledge were also stated by the participants of this study, such as Elif (I1) and Esra (I1). Likewise, Xu and He (2019) expressed that the practicum part was neglected while shaping assessment conceptions of PSTs. Furthermore, DeLuca (2012) stated that education regarding assessment was full of theory and detached from the assessment practices of the daily

life of teachers. Thus, a course on assessment which has components, such as reflection, revision, and practice can be helpful for the assessment literacy of PSTs (Ogan-Bekiroglu & Suzuk, 2014). In this study, more precisely, Elif (I1) expressed that she did not find herself competent regarding in-class assessment and while providing feedback to her students, which Leyla (I1) also agreed upon. Moreover, Leyla (I1) stated that having a lesson focusing on giving feedback to students would be more meaningful during her pre-service education. This might be related to the significance of feedback as one of the key components of the learning process, along with students' self-regulation (Flores et al., 2015). In this respect, Xu and He (2019) implied that PSTs became more aware of the importance of feedback and feedback strategies after starting the practicum part. Moreover, Schellekens et al. (2021) underlined feedback as a key strategy in AfL, as supported by the participants of this study. In addition to in-class practices, Ekin (I2) and Ebru (I2) paid attention to the testing and standardised assessment by indicating that they did not feel fully competent and needed to improve themselves in this regard.

When it comes to reflecting on pre-service assessment courses provided to these instructors based on their departments, participants who graduated from the ELL or Translation & Interpretation Departments said that they, of course, did not have any courses concerning assessment. Esra (I1) stated that she took part in a teaching certificate programme and had courses on assessment there, even though she did not find them suitable and helpful for real-life teaching contexts. While Ekin (I2) mentioned that she has been attending in-service training to compensate for this lack of training, Songül (I3) indicated that it would be better for her to attend assessment-related courses, which she would favour to take with other educational courses in an integrated way. Last but not least, the ELT graduates comprised the largest number of participants. When asked to reflect on their assessment courses, some participants, such as Elif (I1), Leyla (I1), Melda (I2), and Buket (I3), criticised the content of the courses, stating that they were not enough to prepare them for the real assessment contexts, which they encountered right after their graduation. In this vein, Xu and He (2019) suggested that the practicum part could help PSTs encounter authentic teaching environments by making it effective in developing a better understanding of assessment and assessment practices. Moreover, Leyla (I1), Elif (I1), and Melda (I2) also stated that the picture that these courses drew about assessment was not the true reflection of their future real-life experiences. Güngör and Güngör (2024) supported this by indicating that PSTs experience a shock when they encounter the reality of real-life assessment context in MoNE schools, which was due to a lack

of real-life context during their education. Although the participants of this study do not work at MoNE, the same situation is valid in preparatory school contexts when it comes to assessment. Also, while reflecting on their courses on assessment, Ebru (I2) and Hatice (I3), mainly focused on the positive effect of their master's courses rather than the courses they took during their pre-service education. Concerning the fact that not all teachers pursue a master's degree programme, the effectiveness of pre-service courses should also be examined and improved by the decision-makers as it is seen that EFL instructors do not feel well-prepared for assessment during their undergraduate years, regardless of their departments. In this regard, Tsagari and Vogt (2017) also found that the teacher education programmes were insufficient in language teaching and assessment, which may cause teachers to adopt traditional ways of assessment that they were familiar with as students. This was implied by Elif (I1) in this study as a self-criticism since she thought that she used traditional practices concerning in-class assessment. Finally, the participants also highlighted the need to improve themselves as in-service teachers currently, which can also be taken into consideration by policymakers. Regarding this, the participants highlighted the importance of cooperation and in-service training to feel more competent. These suggestions are supported by Xu and Brown (2016) since they argued that teachers with high levels of assessment literacy reflected on their assessment practices and took part in conversations or professional activities regarding assessment, as the participants of this study requested. Moreover, DeLuca and Klinger (2010) addressed the statements of PSTs in Canada who expressed the necessity for instruction concerning both AoL and AfL, which is compatible with the expressions of many participants in this study who reflected on their assessment literacy when they were novice teachers. Therefore, it can be concluded that there is a parallelism between the teaching profession' expecting assessment literate teachers and pre-service or in-service teachers' demand to improve themselves on this topic (DeLuca & Klinger, 2010).

Concerning school accountability conception, the quantitative results showed that the graduates of the ELL Department had the highest mean score ($M=57.55$), followed by the graduates of the Other category ($M=51.63$) and the ELT graduates ($M=48.28$). The distinction is insignificant, meaning that the department from which the participants graduated did not affect their conception regarding school accountability conception.

All in all, the departments that the instructors graduated from do not make profound differences in their assessment conceptions, as the findings of this study suggest. Additionally, the participants in the focus group interview sessions indicated that their pre-service education in

assessment was insufficient in preparing them for real-life contexts and that they felt a need to improve themselves continuously regardless of their educational background. The comments and criticisms of instructors who graduated from the ELT Department should especially be taken into consideration concerning pre-service education since they are expected to become teachers after graduation. Moreover, the effectiveness of the teaching certificate programme (i.e. the Pedagogical Formation Certificate Programme) is questionable and can also be investigated based on the comments of one of the participants in this study.



4.4 RQ2: What are the self-reported assessment practices that instructors apply in their institutions?

The second research question was answered with the data gathered via the qualitative part of the study. Thus, during the focus group interview sessions, the participants were asked what assessment practices they use in their classes and how much they take part in the decision-making process of these assessment practices. While presenting instructors' opinions and perspectives in the framework of the four pre-determined themes, the practices they apply in their classrooms were also mentioned from time to time to provide an in-depth understanding of the topic. However, this research question is merely dedicated to the assessment practices that are applied by the participants.

To begin with, it should be noted that most participants stated that they used assessment practices that were determined by their institutions rather than their individual choices. (e.g. Songül (I3), Buket (I3), Leyla (I1), and others), similar to the findings of Troudi, Coombe, and Al-Hamly (2009) who found that teachers do not take part in the designing process of the assessment tools. Furthermore, even though the participants work at six different universities, the assessment practices they indicated to apply are in line with each other. Firstly, Elif and Esra, who took part in the first interview session, and Ekin, who was included in the second one, have been working at the same university. When asked which assessment practices they use during their lessons, they all mention the institutional practices, which constitute the exams, presentations, and portfolio assessments. To start with, Elif states that her institution assigns many practices throughout the semester. For instance, they require students to have a portfolio that includes speaking and writing tasks that they complete throughout the term. Besides, there are some other practices that she calls performance assessment, such as "Book Cycle."

Institutionally, we have portfolios and some kind of performance assessments. The newest one is the book cycle, for example. [Students] read a book every weekend. At the beginning of the week, we ask them a question, and they reflect on it, both for the improvement of their writing and reading skills. (Elif, I1)

When she considers her own choices concerning assessment, Elif cannot name a variety of things. Although she tries to include peer feedback and observes her students all the time to get a picture of what they know, she does not feel confident about these in-class assessment methods.

[...] When I think of myself individually, how do I assess students in the class? I could not find a lot of things. I go very traditionally, I guess. Since I am not a main course teacher, I do not have any responsibility for doing writing or speaking tasks, so peer feedback for those is also not my duty. However, I use peer feedback very little when somebody answers something. I do not say, "correct" or "incorrect" but I receive feedback from the students. "Is it correct? What do you think? Let us correct our friend and give him/her the right version. Who wants to explain this?" Generally, I have

question-answer sessions. I use all those pen-paper activities after I teach something [such as] vocabulary or grammar. Of course, we have all those exercises. [...] And we are observing, I mean, I am observing everyone constantly. [...] (Elif, I1)

Later on, during the interview, she could remember some of her other preferences. In the excerpt below, Elif mentions including games and group work to assess her students' knowledge.

[...] As we said before, institutions decide on the formal assessment, but I think we use games to evaluate them or some kind of tasks for speaking, especially group work. We [...] teach a concept. After that, "All right, now [you have] two minutes to talk about this. Okay, just talk to your peers; talk about this to your friend." I think we do that a lot in class, and we generally observe who can do it and who cannot. Also, the games, Kahoot especially... I started to use it as a type of assessment to see for myself. Of course, again, [I use Kahoot to see] who can answer the questions, what kind of problems they have, or if it goes very smoothly. [...] It is generally for vocabulary or grammar concepts, but I forgot to say [while answering the previous question.] (Elif, I1)

Working at the same institution, Esra indicates that she prefers using peer feedback in her lessons. Also, she elaborates on the details of the practices applied in their school.

In the class, I use peer feedback. For example, they write a paragraph, and I want the students to give feedback to their classmates. [...] I do not use it in speaking because they can be really shy and frustrated, so I do not use that. [...] (Esra, I1)

We want them to present a topic in the class, and they pick a topic from their books and present it to their friends. We also guide them in this process and give them feedback. [When it comes to portfolio,] their writing and speaking tasks are included in their portfolios and their homework, as far as I remember. (Esra, I1)

Considering her own classroom practices for assessment, she follows a similar path as Elif. Furthermore, she mentions focusing on unit starters, which are the quotations that are given to students for discussion at the beginning of each unit. This may lead her to decide her future practices accordingly while teaching the topics in these units.

I also use online games. For example, I use Quizzes [which is a website], and it shows the questions and their right and wrong answers. If most of the students give wrong answers to a question, you can see, "Okay, you need to focus on this topic." [...] Also, I use speaking sessions like the "unit starters" and I can evaluate what they can say about the topic, and I can have an idea about their speaking skills. I usually do that in the class when I am not responsible for their speaking tasks so that I can understand what we can focus on. (Esra, I1)

Lastly, Ekin goes into more detail about the institutional practices at the same institution by explaining everything they do throughout the school year.

We have portfolios, exams, presentations, and sometimes we have projects. We used to have students shoot mini-TV series. [...] It was something right after the pandemic that we had to get rid of it. This is how we try to do it, and we always try to do it in a standardised way. (Ekin, I2)

While talking about portfolios, she indicates that the writing tasks that students include in their portfolios are process writings, which require students to write many drafts for one writing task. Ekin also mentions that students have to write two drafts, although there used to be three drafts for one writing task in the past. Additionally, it is worth noting that students are given real-life situations for speaking tasks.

Of course, portfolios have some writing and speaking tasks. Writing tasks evolve during the term. For example, [...] we ask questions to describe the process, and it is also a process of writing [written in] two drafts. [...] We give feedback, and the student writes and gives us again. Then we

again give feedback and the grade, and they learn. First, they write something little, and then they write a paragraph. Then, they learn how to write an essay, and from that time, they continue writing several types of essays. In the proficiency exam, they are expected to write a well-developed essay. When it comes to these speaking tasks, they are mostly about real-life issues. They mostly include some role-plays that can be with a partner or alone, and they range from two to four minutes. [...] (Ekin, I2)

She states that students have to give a presentation at the end of each semester. Thus, they are expected to have three different presentations in total. Nevertheless, despite the differences in the topics in each term, students are mainly required to reflect on the process of completing the project throughout the term.

[...] We do different kinds of presentations in each term. In total, they have three presentations because we have three terms in our school. In the first term, we ask our students to choose a unit from the coursebooks as an umbrella term, find something under it, and prepare a presentation about four or five minutes at the end of the term. [...] In the second term, we have a book review process. They are expected to read a book. Then, they are expected to create a digital storytelling presentation. At the end of the term, they are expected to prepare another presentation in which they will tell us about the process. They reflect on the process of reading the book and preparing the digital storytelling. In the last term, we ask them to write blogs, and again, at the end of the term, they reflect on the process of blog writing in their presentation. So, this is mostly about reflection. (Ekin, I2)

When asked how these assessment practices are decided, Ekin points out that many factors are deeply considered during the decision-making process.

We decide on them as the curriculum and testing [units] together. We think about the curriculum; what we are teaching our students, and what we expect them to do as the outcome of this teaching practice, and we look at our coursebook [to see] what kinds of activities we have there. I mean, what are our students familiar with? We decide what to ask based on the objectives, and we decide how to ask based on the tasks that we teach our students. (Ekin, I2)

During the focus group interview sessions, the instructors were also requested to share their opinions on improving the assessment practices that are applied in their institutions. These three instructors underline different points regarding this question. Elif believes that they utilise a range of assessment practices, which is sufficient. However, she thinks that students do not find the preparatory school practices meaningful as they do not believe they will use these things in the future. In a similar vein, most of the students feel a gap between their undergraduate courses and their lessons in the preparatory schools, as suggested by the British Council and TEPAV (2015).

I think we have a lot of different types of assessment, and I do not have anything to say about the types or the variety, but I think there is a blank between the preparatory school and the department. We offer them general English courses, but I think it is not very academic. When [students] go to the next [academic] year, they have some kind of difficulties. Here, the main problem is that we cannot receive any feedback from the faculty members. [...] They need to collaborate more about the needs of the students. Yes, it is not very realistic to teach them ESP or some business English because they come as zero beginners, but we can change the programme and the assessments connectedly, of course, to meet students' academic needs more. We can do that by teaching general English, but we can focus on different types of assessments to be more academic and more related to the department. [Students] could take it more seriously maybe if they see a link [between the preparatory school courses and the courses they will encounter in] the next four years. They will see some assignment types in the preparatory school related to their departments and [we will say,] "Look, you will see an example of this next year, so let us get prepared here." Maybe, that could be

much more meaningful and motivating for them to study harder. [...] They can have a more meaningful purpose to study English. (Elif, I1)

Building upon what Elif says, Esra also believes that students perceive what they do as unnecessary, which, in turn, affects their motivation and engagement in the learning process.

I agree with my colleague Elif because our students think that they do not need what we are teaching right now. For example, I have repeat students. They have only one year to pass [the preparatory school exam], and they do not have any other option. However, they do not have enough hope and motivation because they do not believe what we are teaching is essential for them. They need general English and background information. It is like the basis of English, but maybe they need to see the link between preparatory school and their departments. (Esra, I1)

Ekin, on the other hand, addresses the changes that have already been considered, which might be related to Ekin's job in the testing office. While explaining why these changes have not taken place yet, she refers to the institutional restrictions and requirements.

We do need changes, especially in terms of the length of exam questions, and [exams] have to be more integrative. As my colleague said, we should include more collaboration in it. For example, in speaking exams, we can have some role-plays. We sometimes do it, but sometimes we do not. (Ekin, I2)

We are trying to make some changes, but with the tester teaching twenty-six hours per week, it is not possible, but we are still trying. We are thinking about decreasing the number of questions in our proficiency exam. We have fifty questions in reading and listening parts and forty questions in structure. We think that fifty questions are too many. I mean, the students will get tired, and the success level will drop. [...] In the world, the number of questions is decreasing, and because of that, we are thinking about decreasing the [number of] questions as well, and we are thinking about making it more integrative, maybe by combining listening and writing. [...] (Ekin, I2)

Secondly, Melda from the second interview and Buket, who participated in the third interview, have also been working in the same institution. Melda briefly talks about the institutional practices at their university.

It is the same for us. We do have quizzes and midterms, and we do presentations and speaking assessments in a level-based way as well. Plus, we do have portfolio [...], and at the end, there is a proficiency exam that they need to pass. (Melda, I2)

Just like Ekin, Melda states that assessment practices in their institution are decided based on the curriculum. Furthermore, she underlines that a needs analysis study was also conducted to have a clear idea about the needs of the preparatory school in her institution. Finally, Melda indicates that there are other elements to be considered while deciding assessment practices, along with the curriculum.

Our assessment practises are determined by our curriculum and the departmental needs. At my institution, we did a very large-scale needs analysis study. Then, we changed our proficiency exam. So, every stakeholder in our institution is aware of the departmental needs. Like the other institutions, [our institution] is very big, and we do have a lot of students, like 3000 students, in our preparatory school now. So, some of the practices, unfortunately, are determined by the logistics as well. For example, speaking assessment; we want to do more of it, but we have to do it on a slow scale, not very frequently. This makes the teachers more responsible for evaluating students' speaking skills. So, we have the meso-level university expectations determining our assessment, and we do have the logistics. So, unfortunately, we consider these two. (Melda, I2)

In Melda's case, she thinks that peer assessment and self-assessment will be much more effective for the betterment of the assessment practices in her institution rather than having too

many exams frequently (See page 74). Moreover, during the third interview, Buket, who is a colleague of Melda, elaborates more on the assessment practices that she utilises by indicating that she has to do what is asked of her by the institution.

Because I am working at a preparatory school, I do not have an individual choice, but we have portfolio assessment, for example, or at the end of the term type of assessment; exams, etc. [...] There are also in-class assessments, such as spoken assessment. During the lesson hours, we also do that. [...] [Also,] quizzes and pop quizzes are very famous. [...] (Buket, I3)

[For] writing, students write the first draft, and you give feedback. They write the second draft. [...] They just write one draft, and that is it. (Buket, I3)

Completing her master's thesis on spoken assessment, Buket suggests including assessment of interaction by making spoken assessment less decontextualised in all the preparatory schools in Türkiye.

I do not think anything should be omitted [from the institutional assessment practices] but added. I think we lack assessing interaction in most preparatory schools, not only in our school, because we think this is very difficult. There are lots of things to consider, and we have to be standardised, I understand that; but if we train the raters/instructors and start doing this type of assessment, then we would succeed in this, I think. So, an addition of paired assessment, group assessment, and assessing the interaction is something that needs to be added in most of the preparatory schools in Türkiye, I think. (Buket, I3)

Other instructors' assessment practices are similar to the ones explained above. Leyla also points out exams and writing and speaking tasks in her speech. Also, these writing tasks are process writing, just like other preparatory schools.

I think we do the same. We have a mid-term and a final, which includes a speaking and a writing component. Also, we have the writing tasks that are done as "process writing" [which is a process in which students have to write a few drafts.]. Then we have one presentation at my institution, and that is it. (Leyla, I1)

She also tries to include peer feedback in her lessons, as stated above. Furthermore, when asked what else can be done to improve the assessment practices in her institution, Leyla criticises one of the assessment practices specifically, rather than reflecting on the assessment practices in general.

Especially when it comes to writing, [...] I do peer feedback within the class, but not always because it is very embarrassing for students at some levels, [...] so I try not to leave it to the students perhaps at the start, but towards the end, yes, I do peer feedback. (Leyla, I1)

Moreover, Leyla, if she ever has to decide on the framework of assessment, tries to keep curriculum needs and students' motivation in mind. Ebru refers to all the assessment practices mentioned above regarding her institution. From her statements, we understand that her institution requires students to have writing tasks and presentations, along with exams and standardised practices. In the second excerpt below, Ebru mentions how she chooses her personal practices to assess her students.

We also have midterms and quizzes, of course. They include skills, such as reading, listening, and structure parts, and [the structure part] includes some vocabulary and grammar questions. We do not ask them to write during the quizzes or midterms, but they have to take one writing quiz and one

speaking quiz. Except for these, we also have portfolio assessment, which includes a presentation and writing tasks depending on students' level. This term, I had a beginner-level [class], so we asked them to write short text messages or complaint letters, emails, product reviews, etc. So, they write something, and we provide feedback, and they try to make changes considering the feedback. In total, they have had around ten writing pieces during one term, and they also have to do a presentation in pairs or groups of three. [...] They have to prepare their presentations, and for each part, they receive some feedback on the outline of the presentation. They also have to reflect on the whole process at the end of their presentations. These are the main assessment strategies that we use to decide if a student is passing or failing. (Ebru, I2)

When I assess a student in the classroom, I am not talking about the quiz or a midterm, but in-class assessments, I try to keep the objectives of the lesson in mind and the target language. I mean, target vocabulary, grammar, and speaking tasks. Whatever the objective of the lesson is, I try to assess the students, keeping that in mind. [...] I need to check if the learning is happening or if the learning has happened. (Ebru, I2)

When she was asked to develop ideas to improve assessment practices in her school, Ebru criticised two things. First, she suggests being able to change students' classes/levels in the middle of the term so that they would feel more comfortable for the rest of the year. Besides, her remarks in the second excerpt are roughly in the same vein as Ekin's points (See page 102). Even though Ebru does not explicitly state that the exam should be shorter, she believes that students might get bored during the exam since they cannot leave the exam room.

I wish we had something similar to what Melda said. Sometimes, after the placement test, students are placed in a classroom but do not feel comfortable in that classroom either. It can be because of their higher or lower level. [...] (Ebru, I2)

In our institution, the placement test is too long, and students are not allowed to leave the exam room. They have to stay in the classroom until the end of the exam. If I were a student and did not know any English, I would be so bored. Or if I were a student and if my English was really good, I would be so bored again because it takes hours, and you have to be there, and it is only a written exam. There is no productive part of the exam. Therefore, I think this is the only thing that bothers me [when] I try to see from the students' perspective. [...] (Ebru, I2)

Similarly, Hatice, who works at a different university, mentions some integral components of assessment, such as quizzes, exams, and presentations in her school.

[In my university,] the assessment ways are quizzes [that take place] in two weeks. [Students] have a midterm exam, and two or three quizzes, it depends on their level. [...] Now we are coming to the end of the term, they must do some presentations, and presentation topics will be given to the students, [they will] choose [their topics] from a list. [...] Also, we have exams. (Hatice, I3)

As previously mentioned, Hatice asks her students to reflect on themselves in a way that is different from institutional practices. Thus, she underlines the importance of self-assessment.

I also give importance to the diagnostic form of assessment. We should see spontaneous assessment in a class environment apart from assigning grades. So, at the end of my lessons, twice or once a week, I ask my students to have a blank paper and write what they learned or did not learn during these two weeks, and they anonymously write them. So, I can be aware of their situation. [I mean] whether they have learned or not. They feel less anxious when they do not write their names on these self-assessment forms, so I let them assess themselves: what they have learned and what they would like to learn in that subject before the midterm exam or the [final] exam. (Hatice, I3)

Lastly, Songül agrees with Buket on the topic that they cannot choose their assessment practices individually. Moreover, her stating that they have twenty-four exams in a term was met with surprise by the other participants.

Like Buket already said, we do not have individual choices. In addition to what she said, we have presentations and portfolio items. [Students] read two different books, and we ask only one of them. [...] Other than that, it is just the same with midterms, finals, and quizzes. I am not sure about you, but we have twenty-four exams in one term in total...Twenty-four, yes. However, it is because we have four terms in a year; maybe that is the thing. It is a lot, and we have two quizzes every week, but this kind of keeps [students] alive, and they never leave the vibe of being assessed. They...we never let students feel relaxed, and I guess that is how they want the environment to be in classrooms. (Songül, I3)

She also mentions the details of the presentation preparation process in which students and teachers work hand in hand.

[For presentations,] there is a short list of topics that are related to unit themes or topics, such as [...] environment, "What do you think about electric cars and the environment?" and [students] would come up with an "agree" or "disagree" kind of idea. [...] We just guide them into a smaller circle of topics, and they choose one themselves. They do research [...] and speak about their chosen topics, whether they agree or not. Usually, it is about four or five minutes for B1 levels. For advanced levels, it is even more than that. (Songül, I3)

Since Songül was focused on students' emotions throughout the interview, when she was asked how to improve the assessment practices in her institution, she recommended making assessment practices less stressful for students. Her insight related to making assessment less stressful is expressed as "making it softer" as she calls it (See page 79). This is a general criticism of the way that assessment is perceived in Türkiye alongside her own institution. Besides, her remarks can also be interpreted as making assessment more meaningful for students.

[...] Maybe making [students] choose the topics that they want to talk about would make them a bit less stressed, and it would make them use different kinds of phrases at the same time. For example, if you ask the students, "Do you agree or disagree with something?" they are going to use, "I disagree with the idea that..." or they are going to use some generic statements all the time. If they choose their own topics, it will help them come up with certain ambiguities or phrases that can have a build-up for them. [...] (Songül, I3)

Based on the findings above, it is obvious that all the institutions have exams throughout the term, either quizzes or final exams, and students' writing and speaking skills are also assessed through these exams. Moreover, all the institutions require students to have portfolios, including writing tasks, which are completed in many drafts, and speaking tasks. Lastly, presentations can be seen as an important component in all these institutions, even though the content and length of the presentations change depending on students' level. Therefore, this study found that preparatory schools of universities in Ankara/Türkiye use a wide variety of AfL and AoL tools.

Furthermore, interview questions led participants to think about their individual assessment-related choices as well. To illustrate, Elif (I1) indicated that she used peer feedback, constant observation, games, or group work/peer work in her classes, which might be because developing group management skills of her students, as well as improving interpersonal skills

and reaching a deeper level of learning (Flores et al., 2015). Esra (I1), in the same interview session, agreed with using those practices by mentioning peer feedback and games. Also, she included the use of quotations to ask students' opinions and assess their background knowledge. Moreover, instructors, such as Leyla (I1) or Ebru (I2), also indicated getting help from peer feedback or constant observation in their classes. Although they indicated observing their students all the time, the participants of the present study did not state that informal classroom observations were the most effective assessment methods, unlike Italian teachers (Pastore & Mincu, 2024).

The findings suggest that the instructors and institutions in this study are well aware of the needs of students in the language learning process, and they prefer integrating many different assessment practices. The preferences of institutions and instructors in this study might be concerned with developing students in terms of autonomy, responsibility, and reflection. In addition, methods that are learner-centred lead to a more motivating classroom environment (Flores et al., 2015). Besides, parallel to the findings of this study, Doğru (2020) argued that EFL teachers in higher education were inclined to use assessment with formative purposes, and their preferences for assessment did not alter based on their experience years or undergraduate departments. The findings of this study are also consistent with the study of Gan et al. (2018), who suggested that a wide range of assessment practices were used by Chinese EFL teachers (as cited in Ma & Bui, 2021). The assessment practices of the teachers showed that teachers utilised many different assessment techniques ranging from traditional assessment methods to assessment for improvement purposes, as well as sharing criteria with students and providing feedback. Moreover, as Struyven et al. (2005) stated, portfolios, peer assessment and self-assessment methods, together with some other innovative methods were applied at higher education level, as also indicated by the findings of this study. Nevertheless, even though assessment practices are varied, it should be noted that instructors still have suggestions that can be considered for the betterment of institutional practices, and they criticised some of the existing ones.

4.4.1 External Factors Affecting Assessment Practices

Instructors also addressed some external factors affecting their teaching and assessment while talking about their assessment practices. These external factors mainly involve pre-determined institutional practices that leave no room for teachers to decide on their own practices, the

curriculum of the institution, or time constraints. Concerning this, Ogan-Bekiroglu and Suzuk (2014) point out that teachers are concerned about many things, such as lacking facilities or the curriculum even when they attach importance to alternative assessment methods. Some of these concerns were also mentioned by the participants of the present study. Firstly, some instructors, such as Leyla, Hatice, and Songül, state that they do not have individual choices when it comes to assessment since they need to apply what their institutions ask of them.

I decide on the [assessment practices], if I can decide on anything because we do have institutional practices, but if I can decide, I decide on it by looking at the students' level, enthusiasm, and the needs of curriculum or is there a place in the curriculum for me to do an extra assessment and so on. (Leyla, I1)

Maybe not as an instructor, but it is a departmental issue because sometimes, we can just apply the assessment strategy that the school needs, but I can assess my students in a spoken way. Maybe I can ask them to prepare a presentation, but some institutions just accept the written exam because they need to see a hard copy of the exam. [...] (Hatice, I3)

In addition, Buket states that assessment is not a crucial component for her. Upon being asked what kind of assessment practices she uses, she again states that she does not have individual choices, which makes it necessary to apply what is expected from her as a teacher. The fact that instructors do not have any individual choices can be interpreted in the following way: the assessment practices assigned by the institutions may leave very little time and room for extra individual practices since they are quite time-consuming and happen constantly throughout the term or year.

Assessment is not important to me, but it is important for the institution, so we have to do it so that we can understand if the students would pass or fail and things like that. The only part that is important to me about assessment is that we see what we achieved in the end because we try to teach students something, and assessment is the thing that shows us if they have learned or not. [...] (Buket, I3)

Because I am working at a preparatory school, I do not have an individual choice, but we have portfolio assessment, for example, or at the end of the term type of assessment; exams, etc. [...] (Buket, I3)

Furthermore, as included in the aforementioned sections, Songül indicates that her institution has twenty-four exams in a term. Nevertheless, while expressing this issue, she uses the subject “they” rather than “we” and says, “...*They...we never let students feel relaxed, and I guess that is how they want the environment to be in classrooms.*” It can be deduced from this statement that she alienates herself from this perspective of the institution, though she does not elaborate on that. Also, she overtly says that she does not have individual choices about assessment, just like Buket mentioned.

Like my colleague Buket already said, we do not have individual choices. [...] (Songül, I3)

Moreover, Songül believes that there should be a holistic view of assessment; however, the rubrics prepared by the institution follow a different structure.

I think if the assessment was not based on a rubric that assesses “the positive present simple tense, the negative present simple tense” or “Did they use first of all?” If we did not use that kind of rubric, it would help, maybe because we share the rubrics with the students in my institution and go over them. So, they know what they are supposed to be writing and how to play the game. I think that is one of the problems. [...] (Songül, I3)

On the other hand, Melda was talking about real-life teaching contexts she encountered after her graduation and stated that many elements need to be considered concerning assessment, starting from administrative practices.

[...] When I graduated, assessment was not something like that. There should be a rationale behind it; there is an administration, exams are controlling you, and everything has to be taken into consideration. [...] (Melda, I2)

In another part of her speech, she mentions how institutional facilities affect the assessment practices to be applied. In other words, even though some assessment practices are more favoured by instructors, some school facilities limit those practices.

[...] Some of the practices, unfortunately, are determined by the logistics as well. For example, speaking assessment; we want to do more of it, but we have to do it on a slow scale, not very frequently. This makes the teachers more responsible for evaluating students’ speaking skills. So, we have the meso-level university expectations determining our assessment, and we do have the logistics. So, unfortunately, we consider these two. (Melda, I2)

Ekin’s case is also related to schools’ requirements and policies. As she is responsible for preparing questions and lecturing students at the same time, she complains that her workload makes it difficult for her to improve herself regarding assessment.

I guess I am good at writing structure and vocabulary questions, and I am good at creating contexts for them because this is mostly what I am doing [in my institution. However,] you will not believe it, but I am teaching twenty-six hours per week. So, I cannot write all the questions for all the exams right now. Because of that, I am taking the reading and listening questions from previous exams that someone else wrote years ago. This year, I am doing it and because of that, I cannot improve in terms of writing, reading, and listening questions. [...] (Ekin, I2)

In addition, she highlights some financial issues that need to be solved. It can be inferred that all these problems affect the effectiveness of assessment practices.

[...] Also, practicality is really important for us because during speaking exams or exam weeks, [...] we cannot get help from our part-time teachers. [...] It is because of some stupid issues, I am so sorry. We have to solve these problems in my institution, too. I can say that practicality issues, some financial issues, and some other issues of the office should be solved to assess our student proficiency better. (Ekin, I2)

Apart from the points given above, Elif indicates how time and curriculum restrict teachers. She expresses this since she believes that feedback and reflection are required after exams; however, the need to move on with the curriculum leads her to pay less attention to this part.

[...] I think we do not have enough time. We cannot stop after an assessment for a week, for example, and say, “Okay, you have this, this, and this missing, so let us focus on these for some time. Then we can move on.” That is the ideal, right? But we do not have a chance; no institution has that chance. We need extra time with students, and this is discouraging for both students and teachers. Students are already there for the whole week. If we are talking about preparatory school students, they do not want to stay just one more minute even if they know they need to study. We have to continue with the curriculum and the next units, so time is [...] a challenge. I think everyone can stop for some time and want to reflect on the results, but I think in our context, it is for a very short time. (Elif, I1)

In Esra's statement, time-related issues are also mentioned, just as Elif suggested. Since they need to follow the curriculum, it is difficult to allocate some time for feedback after exams or revise the topics students need to improve further.

[...] But yes, we do not have enough time to focus on what they need because we need to go on, and we cannot go through what they need deeply. (Esra, I1)

In short, this section has illustrated many environmental elements that can influence teachers' assessment decisions and their application processes. These elements range from school facilities to pre-determined institutional practices. Also, time is a problem that is repeatedly mentioned by some participants.



4.5 RQ3: What is the relationship between EFL instructors' assessment conceptions and the self-reported assessment practices they apply?

The findings of the study suggest that instructors mainly have the improvement conception, which was sequentially followed by irrelevance, student accountability, and school accountability conceptions. Furthermore, the in-depth analysis attained through focus group interviews revealed parallel results. In other words, one of the main conceptions that the participants hold in this study is the improvement conception. Therefore, they think that assessment assists them in deciding their future steps by providing information about students' needs and knowledge, which, in turn, helps improve teaching and learning. In this vein, the emphasis they put is primarily on the AfL. When it comes to the assessment practices the participants prefer, their preferences are parallel to the conception of improvement as well. About the assessment practices that can be considered in the framework of improvement conception, firstly, Elif (I1) and Leyla (I1) mainly believed that they needed to be more knowledgeable and implement a variety of assessment practices in their lessons, which they indicated throughout the interview session as a way of self-criticism, similar to the participants of Troudi et al. (2009), who also underlined that multiple methods should be used to assess students so that they can benefit from assessment. The willingness to utilise a range of assessment tools might be related to answering students' different preferences for learning (Flores et al., 2015). Besides, Ebru (I2) comprehensively illustrated how she favoured AfL tools, such as portfolio assessment, giving students time to correct themselves, or self-assessment tools. Likewise, Melda (I2) and Buket (I3) highlighted that they supported the idea of AfL and assessment tools accordingly, which required students to produce the language and turn their focus on learning rather than becoming successful in the exams. In this vein, DeLuca (2012) stated that formative assessment in teaching was associated with many elements, such as better student achievement, more student motivation and students' positive perception of themselves, which might be the underlying reasons for the preferences of the instructors. Moreover, the participants indicated that they adapted their teaching according to the results of assessments, which is in line with the improvement conception and AfL that support the process of learning and students' progress rather than the end product. As explained in the previous question, instructors, such as Esra (I1), Leyla (I1), and Hatice (I3), stated that peer feedback and self-assessment are some of their assessment strategies. Additionally, Elif (I1) and Esra (I1) include group or peer discussions in their in-class assessment practices, along with online quizzes and games, which serve the improvement conception that they hold as well as student

accountability. Also, Elif's point indicating that her institution has a wide range of assessment tools, which she is happy with can be considered in the scope of improvement conception because Brown (2008) mentioned that involving a variety of assessment tools is of utmost importance in improvement conception. Lastly, Ekin's (I2) statements implying that she found speaking and writing tasks very influential on students' learning could also be interpreted in line with the improvement conception.

Secondly, it should be noted that participants of this study do not think that assessment should be omitted, or it should only be based on teachers' intuitive judgements, as Brown (2008) expressed while describing irrelevance conception, even though some participants, such as Elif (I1) and Ebru (I2) mentioned the importance of constant observation as a way of assessing their students, which is aligned with Brown's explanations on irrelevance. Yet, they still think that different assessment practices should be applied in the language learning process. Therefore, some of the criticisms of participants concerning the institutional and national assessment practices will be provided in this part since their extensive criticism might be related to perceiving those practices as irrelevant.

Firstly, the participants criticise and do not find some institutional practices effective because some assessment practices are considered very traditional, as explained in the form of criticism by the participants and can be discussed in the framework of irrelevance conception. In addition, strict rubrics while assessing speaking and writing were indicated as inauthentic, decontextualised, and ineffective by the participants, such as Buket (I3) and Songül (I3). Taking these criticisms and suggestions into consideration, oral exams can be designed to be better opportunities to develop critical thinking skills of the students, as mentioned by Baird (2010). Similarly, it is indicated by Struyven et al. (2005) that students find assessment more effective when assessment is authentic reflecting real-life contexts. Moreover, Hatice (I3) expressed her willingness to have more room to make assessment-related decisions as a teacher individually since the practices assigned by her school do not give her this opportunity. Lastly, Ekin (I2) considers some exam procedures ineffective because of the problems experienced in her institution, which hinders her improvement as an item writer in the testing office. Hence, she expects these problems to be solved so that assessment practices can be more effective. Some other studies in the literature addressed the discrepancy between teachers' desired assessment context and the reality of the existing conditions, as included in the meta-analysis study of Azis (2012). S/he stated (2012) that an assessment conceptions study conducted in the UK with two hundred and twenty participants by Winterbottom et al. (2008) revealed that even though trainee

teachers favoured assessment for improvement purposes, the educational policies of the system impeded their preferences since these national policies were needed to be obeyed and followed by the teachers. There were also a few instances indicating the same issue in the present study. Melda (I2) stated that she preferred AfL tools even though her institution had a lot of quizzes and exams. Besides, Songül (I3) mentioned that assessment should cause less stress and that students should be given more space and evaluated holistically, although her institution follows opposite paths. Buket (I3) also criticised some of her institutional practices, such as ineffective assessment of speaking and listening skills, which take place in a decontextualised way, as Buket (I3) called. As a result, the participants' holding irrelevance conception is reflected in their criticisms concerning assessment practices. These criticisms should be taken into consideration and be tackled attentively by the decision-makers because the problems experienced by the teacher may lead them to perceive assessment as irrelevant, although this is not the case right now.

Thirdly, some of the preferences of assessment practices of the participants can be evaluated under the scope of student accountability conception. Although this is the third conception that instructors hold in this study, it may go hand in hand with the improvement conception. As Brown (2008) also stated, due to the emphasis on self-assessment, peer assessment, and student-centred educational view, teachers may have student accountability conception, which was seen in this study. Participants' willingness to include AfL tools, such as self-assessment and peer assessment methods, along with their willingness to provide feedback to their students in the learning process, are parallel with their holding student accountability conception, as mentioned and illustrated while answering previous questions (See page 74).

The least agreed conception in this study is the school accountability conception. Since school accountability conception is about teachers' and schools' publicly showing that they have done a good job, the ones holding this conception might be in favour of high-stakes exams as a way to demonstrate their success, leading them to turn their focus on exams and exam scores. Nevertheless, the participants in this study do not prefer exams as the main means of assessment, and they believe that exams should be supported with a variety of other assessment practices in line with AfL. Therefore, there is a parallelism between their small agreement level with school accountability conception and their preferences for assessment practices. In other words, participants hold school accountability conception to a limited extent. Similarly, they do not conceive exams as the most important assessment practice. Moreover, the participants in the focus group interview sessions did not explicitly state that they needed exams or that they

needed to show the success of their institutions or their own success as teachers publicly. However, some of their remarks, such as “*Do you have quality teachers?*” stated by Ekin (I2) might be interpreted as holding teachers accountable for the success of students and holding this conception to some extent. Besides, Buket’s (I3) negative feelings attached to being accountable in assessment as a teacher might also stem from the pressure about having school accountability conception since teachers have to show the results of their teaching to third parties publicly. As presented above, she said, “...*But, if there were no exams and things like that, we would be more relaxed. I think it would be even... This might be bold, but we would be more...The students would feel happier to learn in the end, maybe.*” even though the other participants did not have negative opinions towards assessment in this strict manner. In this vein, Pastore and Mincu (2024) highlighted that Italian teachers found it problematic to share the assessment results of their students with stakeholders, which may be related to being afraid of other people’s judgements stemming from their lack of confidence.

All in all, the relationship between the conceptions that the instructors have and their preferences for assessment practices is significant, and there is a positive correlation between them, as also found by Brown et al. (2009). They stated that there was a parallelism between primary and secondary school teachers’ conceptions and their assessment practices. Similarly, the participants in this study mainly hold an improvement conception, which manifests itself in the idea of AfL practices emphasising the process of language learning and improvement of the teaching and learning process. Besides, irrelevance conception has the second highest mean score in this study, which might result from the gap between participants’ ideal assessment context and real-life assessment context in their country and institutions, just as Azis (2015) implied. Thus, instructors’ opinions can be valued and considered to change their tendencies from perceiving assessment as irrelevant to assessment for improvement purposes. In line with the TCoA-IIIA results, some traces of student accountability conception and some practices related to this conception can be found since instructors prefer making students accountable in their own language learning journey even though they are aware of the stress that students feel concerning being responsible for their performances. Lastly, school accountability conception has the lowest mean score, which is, again, parallel to instructors’ negative feelings concerning exams and high-stakes assessments.

4.6 Summary

In this chapter, each research question was answered by presenting the findings and discussed by relating them with the findings of the other studies in the literature. The following chapter will present the conclusion based on the findings and discussion parts.



CHAPTER V

CONCLUSION

5.1 Summary of the Study

In this study, EFL instructors' assessment conceptions were investigated using the framework of four assessment conceptions defined by Brown (2002, 2004, 2008). To this end, instructors working at six different preparatory schools in Ankara/Türkiye were chosen through convenience sampling to complete the TCoA-IIIA. After one hundred and one EFL instructors answered the scale, the instructors, who left their e-mail addresses for the interview, were contacted for the focus group interview part to hear their opinions on assessment and assessment practices, and nine instructors chose the appropriate time slots for them. Subsequently, three different focus group interview sessions were completed with three instructors in each one after the pilot study. Upon collecting the data, the quantitative part of the study was analysed through ANOVA and the Kruskal-Wallis H test after the normality analysis. On the other hand, the qualitative part was analysed by qualitative content analysis via coding (Creswell, 2009) and placing the codes under the pre-determined conceptions defined by Brown (2008) with the help of MAXQDA. Moreover, throughout the data collection, analysis, and reporting processes, the researcher included her opinions in her self-reflection journal to improve the process and be transparent while reporting the results.

The findings of the study were presented and discussed in Chapter IV by relating the findings to each research question. The results demonstrate that EFL instructors mainly have the improvement conception, followed by the irrelevance conception. These findings indicate instructors' belief that assessment helps improve the learning and teaching process, which was supported by their statements during the interview sessions. Moreover, irrelevance conceptions' having one of the highest mean values may mean that there are some opinions of the instructors that should be taken into consideration. For instance, almost all the participants mentioned that they either felt a need to improve themselves on assessment when they were recent graduates, or they still feel insufficient regarding knowledge of some assessment practices. The least

agreed conceptions were student accountability and school accountability conceptions in this study. Additionally, it was found that the years of experience of the instructors and their undergraduate departments did not make much of a difference in their conceptions. Lastly, the instructors participating in the focus group interview sessions pointed out that they had few individual choices concerning assessment apart from the assessment practices previously determined by their institutions. Although this was the case, it was evident that instructors were pleased with the variety of practices applied in their institutions, with some suggestions to improve those assessment practices even further. Their statements during interview sessions showed that their conceptions were reflected in the assessment practices they prefer. In other words, the instructors in this study favoured AfL tools, and they were concerned about the pressure and stress that students feel concerning AoL practices, such as exams.

5.2 Implications

5.2.1 Practical Implications

The results of the study offer some implications for future practices concerning assessment. Firstly, concerning the importance of context and content in assessment courses, pre-service teacher education regarding assessment can be improved by offering more practical content/knowledge that can be applied in real-life contexts in ELT programmes (DeLuca, 2012; Ogan-Bekiroglu & Suzuk, 2014). This may be achieved by changing the content of those courses or by integrating assessment implementation as a practicum component, together with the knowledge base (Xu & Brown, 2016). In a similar vein, Güngör and Güngör (2024) indicate that there are many factors addressing the lack of assessment competence of PSTs. These are listed as limited education in assessment, the insufficiency of involving research and context, and the contrast between theory and practice parts. Moreover, DeLuca, Chavez, Bellara, and Cao (2013) underline the importance of both research and practice parts concerning teacher training programmes since research is also required to keep up with the contemporary assessment practices. In the present study, participants also recommend some topics that can be elaborated on during pre-service education on assessment, such as conducting in-class assessments and providing feedback to students, along with profound knowledge on designing and preparing quality exams. These further steps should be taken since EFL instructors start their teaching careers with inaccurate expectations and negative feelings towards assessment practices, as suggested by the participants of the present study.

Secondly, as argued by Güngör and Güngör (2021), pre-service education should be followed by in-service teacher training or workshops on assessment based on the needs of teachers and requirements of the time (e.g. the need for different assessment practices during the COVID-19 period). These should be supported through cooperation with colleagues, which might lead to a decrease in the level of irrelevance conception of assessment. As Maclellan (2004) argues, novice teachers' knowledge of assessment needs to be improved, which can be achieved by exposing them to the content knowledge regarding assessment with the help of reading materials and discussions. Therefore, institutions can initiate training to help novice EFL instructors as well as foster the knowledge and practices of all the instructors in this regard. Also, in a relatively recent research study, Schellekens et al. (2021) suggest investing in teacher training programmes as well as student intervention programmes to make them assessment literate. In this way, teachers can feel more competent in guiding and supporting their students by integrating student-centred approaches. Besides, it is suggested in the report of the British Council and TEPAV (2015) that continuing professional development programmes should be conducted to support teachers' development by considering their needs. As the present study demonstrates, organising continuing professional development or teacher training programmes coupled with the instructors' willingness to take part in those programmes to achieve their intended assessment literacy and competency level might lead to effective results.

Thirdly, decision-makers can consider shifting focus from AoL to AfL by including more meaningful and less stressful assessment practices, as offered by the participants. As Kippers, Wolterinck, Schildkamp, Poortman, and Visscher (2018) discuss, AfL is about sharing learning objectives and criteria with students, getting information about their learning process, and using this information as feedback for the improvement of student learning. In this regard, peer assessment and self-assessment methods are used daily, as also favoured by the participants of this study. In this vein, Slater (1996) includes students' positive perceptions on portfolio assessment by indicating that portfolio assessment leads them to internalise the knowledge and remember it easily compared to other assessment methods. In addition, the participants suggest changing the assessment of writing, listening, and speaking skills by assessing interaction and making it similar to real-life contexts or through integrating a holistic view of assessment and rubrics accordingly by highlighting the importance of communicative approach in testing (Fulcher & Davidson, 2007).

Furthermore, designing and preparing high-quality exams, together with achieving standardisation in grading speaking and writing skills should be tackled with more caution as

discrepancies in these steps might discourage both students and teachers and might lead to perceiving assessment as irrelevant or unnecessary. In this vein, Kippers et al. (2018) argue that assessment should be of high quality for the systematic analysis and interpretation of assessment results, which can help adjust the teaching for improvement purposes. Also, the report of the British Council and TEPAV (2015) advocate that concerns regarding quality of assessment should be addressed at the tertiary level.

Lastly, some institutional problems should be solved to apply assessment practices effectively and attain better results from those practices. These institutional problems range from the number of students and instructors to the financial issues resulting from not paying enough money for part-time instructors, as listed by Arda (2023). This, in turn, leads to not being able to get help from these part-time teachers in some institutions. Therefore, the problems that the instructors face and their recommendations for improving assessment practices can be requested. Even this study sheds light on some institutional needs and problems. Thus, more comprehensive needs analysis studies can be conducted to gain a deeper understanding and to find effective solutions institutionally.

5.2.2 Research Implications

Based on the study's limitations, this study may address some implications for further research studies. First, this study aimed to reach more participants in the quantitative data collection process; however, the number of participants was limited to one hundred and one instructors. In other words, the number of participants was small concerning departments and years of experience. Thus, some options were merged to have more meaningful results while analysing the data because of this limited number. This factor can be considered while conducting further research studies so that an equal distribution of participants can be ensured.

Second, participants in this study are working at six different preparatory school programmes in Ankara/Türkiye. More preparatory schools can be included in future research studies so that conceptions of instructors working at different universities can be understood thoroughly. This can draw a wider picture of the assessment conceptions of EFL instructors in the Turkish context.

Third, participants of the study expressed their opinions on assessment and assessment practices. Nevertheless, the lessons and assessment practices of these participants were not observed. Thus, in-class observations and actual classroom research can be incorporated into

further research studies to understand the alignment of instructors' conceptions to their real-life assessment practices in a better way, which would support the interview sessions in the data collection stage (Tsagari & Vogt, 2017) by enhancing credibility.

Fourth, this study investigated the assessment conceptions of EFL instructors working at the preparatory schools. Considering the significant role of teachers in assessment (Öz & Atay, 2017), different assessment conceptions studies can be carried out to gain a comprehensive understanding of the English language teachers' perceptions working at the MoNE, along with studies investigating the assessment conceptions of PSTs in addition to the studies in the literature (DeLuca & Klinger, 2010; Levi & Inbar-Lourie, 2020; Lutovac & Flores, 2022). The further steps of decision-makers can be situated on a strong basis with more practical implications in that way.

Additionally, there was a participant who attended the teaching certificate programme (i.e. the Pedagogical Formation Certificate Programme), which is offered to the graduates of the ELL, Translation & Interpretation, Linguistics, or American Language & Literature Departments to become English teachers in Türkiye. The participant offered her insights and criticisms concerning the teacher training certificate programme and the courses on assessment, suggesting that it should be improved, similar to the findings of Önder and Tagay (2015). Therefore, a comprehensive study can be conducted to understand the effectiveness of this teaching certificate programme, which is accepted by all the schools in Türkiye.

Also, some training can be offered to instructors to improve their literacy and confidence in assessment. After the training, the effectiveness of the training in shaping assessment competencies of instructors, along with their conceptions can be investigated through the reflections of instructors so that successful implementations can be augmented and adopted in similar contexts, similar to the studies (Fabriz et al., 2021; Looney, Cumming, van Der Kleij & Harris, 2018). For instance, Mertler (2009) carried out workshop on classroom assessment. S/he found that in-service teachers' literacy was higher compared to pre-test results. As a result, teachers stated the positive effect of the training.

Finally, a study concerning preparatory school students' conceptions of assessment in the Turkish context can be carried out to understand the reflections of tertiary-level students on assessment and assessment practices in their institutions for further examination of the topic, improvement of the current assessment practices and providing further findings to the literature. As discovered by Struyven et al. (2005), there is a link between students' perceptions of assessment and their learning approach. Hence, their opinions can be a guidance for the

improvement of learning. Moreover, it is required to investigate the assessment psychology of higher education students, which may yield different results from the students studying at compulsory levels (Brown, 2022).

In conclusion, the practical and research implications drawn from the findings of this study highlighted some significant points in the field. Mainly, the need for improving courses offered to PSTs and providing in-service training for EFL instructors on assessment have been underscored concerning practical implications, along with some other implications to improve existing assessment practices and make them more meaningful for instructors. Furthermore, research implications address some areas for further research studies so that decision-makers can have a stronger basis for their future steps regarding enhancing the assessment practices and diminishing the level of consensus on the irrelevance conception.

REFERENCES

- Al-Nouh, N. A., Taqi, H. A., & Abdul-Kareem, M. M. (2014). EFL primary school teachers' attitudes, knowledge and skills in alternative assessment. *International Education Studies*, 7(5), 68-84.
- Ali, M. (2011). *Teachers' and students' perspectives on English language assessment in the secondary English Language Teaching (ELT) curriculum in Bangladesh* (Master's thesis). Retrieved from <https://ir.canterbury.ac.nz/server/api/core/bitstreams/764aa123-5090-4978-9773-2262398023a0/content>
- Arda, Ç. (2023). *Part-time English language instructors' well-being: A case study in Türkiye* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/>
- Azis, A. (2012). Teachers' conceptions and use of assessment in student learning. *Indonesian Journal of Applied Linguistics*, 2(1), 41-52.
- Azis, A. (2015). Conceptions and practices of assessment: A case of teachers representing improvement conception. *TEFLIN Journal*, 26(2), 129-154.
- Baird, J. A. (2010). Beliefs and practice in teacher assessment. *Assessment in Education: Principles, Policy & Practice*, 17(1), 1-5.
- Barnes, N., Fives, H., & Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107-116.
- Bayram, İ. & Canaran, Ö. (2019). Evaluation of an English preparatory program at a Turkish foundation university. *Journal of Language and Linguistic Studies*, 15(1), 48-69.
- Berry, R. (2008). *Assessment for learning*. Hong Kong: Hong Kong University.
- Betoret, F. D., & Gómez-Artiga, A. (2004). Trainee teachers' conceptions of teaching and learning, classroom layout and exam design. *Educational Studies*, 30(4), 355-372.

- Black, J. H., Hall J., Martin S., & Yates, J. (1989). The quality of assessments: Case studies in the National Certificate. *Scottish Council for Research in Education*.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575.
- Brinkmann, S. (2014). Unstructured and semi-structured interviewing. In P. Leavy (Ed.), *The Oxford handbook of qualitative research* (pp. 277-299). Oxford: Oxford University.
- British Council & TEPAV (2015). *The State of English in Higher Education in Turkey*. Ankara. Retrieved from <https://www.britishcouncil.org.tr/>
- Brown, G. T. L. (2002). *Teachers' conceptions of assessment* (Doctoral dissertation). Retrieved from <https://researchspace.auckland.ac.nz/handle/2292/63>
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson Education.
- Brown, G. T. L. (2004) Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99(1), 166-170.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science.
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, 16(3), 347-363.
- Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50, 307-320.

- Brown, G., & Harris, L. (2012). Student conceptions of assessment by level of schooling: Further evidence for ecological rationality in belief systems. *Australian Journal of Educational & Developmental Psychology*, 12, 46-59.
- Brown, G. T. L. (2019). Is assessment for learning really assessment? *Frontiers in Education*, 4, 1-7.
- Brown, G. T. L. (2022). Student conceptions of assessment: Regulatory responses to our practices. *ECNU Review of Education*, 5(1), 116–139.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389.
- Comforto, J. (2022). *Teacher assessment conceptions and practices at an independent school* (Capstone project). Retrieved from <https://ir.vanderbilt.edu/handle/1803/17680>
- Council of Higher Education (2016). [Yükseköğretim Kurumlarında Yabancı Dil Öğretimi Ve Yabancı Dille Öğretim Yapılmasında Uyulacak Esaslara İlişkin Yönetmelik]. Official Gazette, Issue: 29662.
- Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches*. The United States of America: SAGE.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. The United States of America: SAGE.
- Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research*. The United States of America: Pearson Education.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. The United States of America: SAGE.
- Çakır, C. (2013). Standard assessment and alternative assessment in English language teaching program. *Gazi University Journal of Gazi Educational Faculty*, 33(3), 531-548.
- Çimen, Ş. S. (2022). Exploring EFL assessment in Turkey: Curriculum and teacher practices. *International Online Journal of Education and Teaching (IOJET)*, 9(1), 531-550.
- Delett, J. S., Barnhardt, S., & Kevorkian, J. A. (2008). A framework for portfolio assessment in the foreign language classroom. *Foreign Language Annals*, 34(6), 559-568.

- 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.
- DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, 34(5), 576-591.
- DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator*, 48(2), 128-142.
- Dikli, S. (2003). Assessment at a distance: Traditional vs alternative assessments. *The Turkish Online Journal of Educational Technology*, 2(3), 13-19.
- Dixson, D. D., & Worrell, F. C. (2016) Formative and summative assessment in the classroom. *Theory Into Practice*, 55(2), 153-159.
- Doğru, D. (2020). *Assessment literacy conceptions and practices of EFL teachers in higher education* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/>
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University.
- Elliott, S. N. (1995). Creating meaningful performance assessments. *ERIC Digest*, ED381985. Retrieved from <https://eric.ed.gov/?id=ED381985>
- Fabriz, S., Hansen, M., Heckmann, C., Mordel, J., Mendzheritskayaa, J., Stehle, S., Schulze-Vorberg, L., Ulrich, I., & Horz, H. (2021). How a professional development programme for university teachers impacts their teaching-related self-efficacy, self-concept, and subjective knowledge. *Higher Education Research & Development*, 40(4), 738–752.
- Figlio, D. N., & Ladd, H. F. (2015). School accountability and student achievement. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of research in education finance and policy* (pp. 194-210). New York: Routledge.
- Fives, H., & Buehl, M. M. (2014). Exploring differences in practicing teachers' valuing of pedagogical knowledge based on teaching ability beliefs. *Journal of Teacher Education*, 65(5), 435-448.

- Flores, M. A., Simão, A. M. V., Barros, A., Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: A study in higher education. *Studies in Higher Education*, 40(9), 1523–1534.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. New York: Routledge.
- Gebril, A. (2016). Language teachers' conceptions of assessment: An Egyptian perspective, *Teacher Development*, 21(1), 81-100.
- Gibbs, A. (2012). Focus groups and group interviews. In J. Arthur, M. Waring, R. Coe & L. V. Hedges (Eds.), *Research methods & methodologies in education* (pp. 186-192). London: SAGE.
- Gottlieb, M. (2016). *Assessing English language learners: Bridges to educational equity*. Thousand Oaks: Corwin.
- Greenstein, L. (2010). *What teachers really need to know about formative assessment*. The United States of America: ASCD.
- Gülbahar, Y., & Büyüköztürk, Ş. (2008). Adaptation of assessment preferences inventory to Turkish. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35, 148-161.
- Güngör, M. N., & Güngör, M. A. (2021). Reconceptualizing testing in times of COVID-19 from an activity theory lens: A narrative inquiry of a test-designer's experiences. *The Asia-Pacific Education Researcher*, 30(6), 531-539.
- Güngör, M. A., & Güngör, M. N. (2024). Relocating assessment in pre-service teacher education: An emerging model from activity theory lens. *European Journal of Teacher Education*, 1-21.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365-379.
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365-381.

- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169–207.
- Hatipoğlu, Ç., & Erçetin Koca, N. G. (2014). Türkiye’de yabancı dilde ölçme ve değerlendirme eğitiminin dünü ve bugünü ve yarını. In S. Akcan & Y. Bayyurt (Eds.), *Proceedings of the 3. Ulusal Yabancı Dil Eğitimi Kurultayı* (pp.72-89). Turkey: Boğaziçi Üniversitesi.
- Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *International Association of Research in Foreign Language Education and Applied Linguistics ELT Research Journal*, 4(2), 111-128.
- Huitt, W., Hummel, J., & Kaeck, D. (2001). Assessment, measurement, evaluation, and research. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved from <http://www.edpsycinteractive.org/topics/intro/sciknow.html>
- Huitt, W. (2007). Assessment, measurement, and evaluation: Overview. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved from <http://www.edpsycinteractive.org/topics/measeval/msevlovr.html>
- Hung, C. S., & Wu, H. K. (2023). High school science teachers ‘assessment literacy for inquiry based science instruction. *International Journal of Science Education*, 1-22. Retrieved from <https://www.researchgate.net/>
- Ivankova, N. V., Creswell J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(3), 3-20.
- Kane, T. J., & Staiger, D. O. (2001). Improving school accountability measures. NBER Working Paper Series, No. 8156. Retrieved from <https://www.nber.org/>
- Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research*, 93(5), 276-286.
- Khodashenas, M. R., Khodabakhshzadeh, H., Baghaei, P., & Motallebzadeh, K. (2022). EFL teachers assessment literacy needs inventory: A case of Fulcher’s assessment literacy framework. *Issues In Language Teaching*, 11(1), 131-156.
- Kılıç, S. (2016). Cronbachs alpha reliability coefficient. *Journal of Mood Disorders*, 6(1), 47-48.

- Kırkgöz, Y. (2007). English language teaching in Turkey: Policy changes and implementations. *RELC Journal*, 38(2), 216-228.
- Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, 75, 199-213.
- Kitchen, H., Bethell, G., Fordham, E., Henderson, K., & Li, R. R. (2019). *OECD Reviews of Evaluation and Assessment in Education: Student Assessment in Turkey*. Paris: OECD.
- Kizlik, B. (2019). Measurement, assessment, and evaluation in education. *Adprima Educational Sources*. Retrieved from <http://www.adprima.com/measurement.htm>
- Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice*, 16(3), 1-6. Retrieved from <https://eprints.qut.edu.au/28741/>
- Kyttälä, M., Björn, P. M., Rantamäki, M., Lehesvuori, S., Närhi, V., Aro, M., & Lerkkanen, M. K. (2022). Assessment conceptions of Finnish pre-service teachers. *European Journal of Teacher Education*, 1-19.
- Lam, R. (2015). Assessment as learning: Examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. *Studies in Higher Education*, 41(11), 1900-1917.
- Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. London: RoutledgeFalmer.
- Lee, I. (2011). Bringing innovation to EFL writing through a focus on assessment for learning. *Innovation in Language Learning and Teaching*, 5(1), 19-33.
- Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168-182.
- Levin, B. B. (2015). The development of teachers' beliefs. In H. Fives & M. G. Gill (Eds.), *International Handbook of Research on Teachers' Beliefs* (pp.48-65). New York: Routledge.

- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 1-26.
- Lutovac, S., & Flores, M. A. (2022). Conceptions of assessment in pre-service teachers' narratives of students' failure. *Cambridge Journal of Education*, 52(1), 55-71.
- Ma, M., & Bui, G. (2021). Chinese secondary school teachers' conceptions of L2 assessment: A mixed-methods study. *Studies in Second Language Learning and Teaching*, 11(3), 445-472.
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, 20, 523-535.
- Majid, M. A. A., Othman, M., Mohamad, S. F., Lim, S. A. H., & Yusof, A. (2017). Piloting for interviews in qualitative research: Operationalization and lessons learnt. *International Journal of Academic Research in Business and Social Sciences*, 7(4), 1073- 1080.
- Malato, G. (2023). *An introduction to the Shapiro-wilk test for normality*. Retrieved from <https://builtin.com/data-science/shapiro-wilk-test>
- McKay, S. L. (2006). *Researching second language classrooms*. New Jersey: Lawrence Erlbaum Associates.
- McNamara, T. (2001). Rethinking alternative assessment. *Language Testing*, 18(4), 329-332.
- Meral, E. (2022). The role of conceptual change texts in concept teaching: Active citizenship learning space. *Türk Akademik Yayınlar Dergisi*, 6(2), 146-174.
- Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus. Retrieved from <https://eric.ed.gov/?id=ED482277>
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12, 101–113.
- Ministry of National Education. (2018). *English Curriculum for 2,3,4,5,6,7, and 8th Grades*. Retrieved from <https://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=327>

- Mislevy, R. J. (2017). On measurement in educational assessment. In C. Secolsky & D. B. Denison (Eds.), *Handbook on measurement, assessment, and evaluation in higher education* (pp. 11-31). New York: Routledge.
- Moiinvaziri, M. (2015). University teachers' conception of assessment: A structural equation modeling approach. *Journal of Language, Linguistics and Literature*, 1(3), 75-85.
- Monteiro, V., Mata, L., & Santos, N. N. (2021). Assessment conceptions and practices: Perspectives of primary school teachers and students. *Frontiers in Education*, 6, 1-15.
- Nasab, F. G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, 2(6), 165-178.
- Neibling, J. L. (2014). *Teachers' conceptions toward type of assessment: Grade level and state tested content area* (Master's thesis). Retrieved from <https://core.ac.uk/download/pdf/213405832.pdf>
- Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments*. University Of Hawaii.
- Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *The Curriculum Journal*, 25(3), 344-371.
- Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report Volume*, 13(4), 695-705.
- Otrar, M. (n.d). *Tek yönlü varyans analizi (ANOVA)*. Retrieved from <https://mustafaotrar.net/istatistik/tek-yonlu-varyans-analizi-anova/>
- Önalan, 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.
- Önder, E., & Tagay, Ö. (2015). Opinions of candidate teacher in the pedagogical formation certificate program with regard to the teacher profession and the certificate program. *International Journal of Human Sciences*, 12(1), 293-314.
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.

- Pastore, S. (2023a). Italian teachers' conceptions of large-scale assessment: What are the implications for assessment literacy? *Teacher Development*, 27(2), 153-171.
- Pastore, S. (2023b). Teacher assessment literacy: A systematic review. *Frontiers in Education*, 8, 1-25.
- Pastore, S., & Mincu, M. (2024). A feasible balance? The Italian teachers' standpoint on assessment literacy, assessment practice, and teacher professional development. *European Journal of Teacher Education*, 1-20.
- Phongsirikul, M. (2018). Traditional and alternative assessments in ELT: Students' and teachers' perceptions. *rEFlections*, 25(1), 61-84.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4-11.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46, 265–273.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100, 190-208.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27, 472-482.
- Saldaña, J. (2011). *Fundamentals of qualitative research*. New York: Oxford University.
- Saldaña, J. (2013). *The coding manual for qualitative researchers*. Bodmin, Cornwall: SAGE.
- Schellekens, L. H., Bok, H. G. J., Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of assessment as learning (AaL), assessment for learning (AfL), and assessment of learning (AoL). *Studies in Educational Evaluation*, 71, 1-15.
- Shute, V., & Kim, Y. J. (2014). Formative and stealth assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 311-321). Springer, New York, NY.
- Slater, T. F. (1996). Portfolio assessment strategies for grading first-year university physics students in the USA, *Physics Education*, 31(5), 329–333.

- Stiggins R. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan*, 69(5), 363-368.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 86(1), 22-7.
- Stiggins, R. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.
- Struyven, K., Dochy F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.
- Swartling, Å. G. (2007). Focus groups. In *Advanced Tools for Sustainability Assessment, European Commission Webbook*. Retrieved from <https://www.researchgate.net/>
- Taherdoost, H. (2016). Sampling methods in research methodology; How to choose a sampling technique for research. *International Journal of Academic Research in Management (IJARM)*, 5(2), 18-27.
- Teddlie, C., & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (pp. 1-42). The United States of America: SAGE.
- Toner, J. (2009). Small is not too small. *Qualitative Social Work*, 8(2), 179-192.
- Tosuncuoğlu, İ. (2018). Importance of assessment in ELT. *Journal of Education and Training Studies*, 6(9), 163-167.
- Troudi, S., Coombe, C., Al-Hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555.
- 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.
- Ulum, Ö. G. (2015). History of EFL teacher education programs in Turkey. *Journal of Research in Humanities and Social Science*, 3(7), 42-45.

- Vardar, E. (2010). *Sixth, seventh and eighth grade teachers' conception of assessment* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/>
- Watt, D. (2007). On becoming a qualitative researcher: The value of reflexivity. *The Qualitative Report*, 12(1), 82-101.
- Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools?* The United States of America: RoutledgeFalmer.
- Wickline, L. E. (1971). Educational accountability. In E. W. Roberson (Ed.), *Educational accountability through evaluation* (pp.7-18). The United States of America: Educational Technology.
- Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, 45(2), 107–122.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.
- Xu, Y., & He, L. (2019). How pre-service teachers' conceptions of assessment change over practicum: Implications for teacher assessment literacy. *Frontiers in Education*, 4, 1-16.
- Yaman, İ., & Şahin, M. (2019). An Investigation into the Employment of English Language Teachers in the Private Sector in Turkey. *Journal of Language Research (JLR)*, 3(1), 1-13.
- Yan, X., & Fan, J. (2021). “Am I qualified to be a language tester?”: Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246.
- Yan, X., Zhang, C., & Fan, J. J. (2018). Assessment knowledge is important, but...: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158–168.
- Yetkin, R. (2017). *Pre-service English teachers' conception of assessment and their future assessment practices in a Turkish context* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/>
- Yetkin, R., & Özer, Z. (2020). An investigation into in-service and pre-service English teachers' conception of assessment. *Journal of Theoretical Educational Science*, 13(2), 384-396.

Yüce, Z. (2015). *Pre-service English language teachers' conceptions of assessment and assessment practices* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/>

Zaimoğlu, S. (2013). *Teachers' and students' conceptions of assessment in a university EFL preparatory school context* (Master's thesis). Retrieved from <https://tez.yok.gov.tr/>



APPENDICES



Appendix 1. Information & Consent Form

Dear participant,

This study is being conducted by İrem Aksu, a master's student at Gazi University-English Language Teaching Department, under the supervision of Assoc. Prof. Dr. Müzeyyen Nazlı Güngör to comprehend the assessment conceptions of in-service English language instructors as well as the assessment techniques that are used by them.

In this study, first of all, your teaching experience and your educational background are going to be asked. A questionnaire, which includes twenty-seven items designed to understand the assessment conceptions of teachers, will follow that.

In the second step of the study, some of the participants -if they are willing to participate- are going to be chosen to participate in a focus group interview in order for the researcher to have a more detailed framework of their assessment practices. Besides, the participants are going to be asked to share their experiences and examples from their own teaching experiences on the related concept. If you want to take part in the second step, you can indicate that in the related part.

Your name, surname or other personal details will not be shared with third parties. The data that you provide as well as the recording of the interview will only be used by the researcher in the scope of this study. **You can contact me via [1](#) in case of any questions/problems.**

By moving on, you agree to participate in the first part of the study voluntarily. Thanks a million for your participation and support!

İrem Aksu

Appendix 2. Demographic Information & TCoA-IIIA

2.1 Demographic Information

1- How long have you been teaching?

- ☐ Less than 2 years
- ☐ Between 2-5 years
- ☐ Between 5-10 years
- ☐ More than 10 years

2- Which department did you graduate from?

- ☐ English Language Teaching
- ☐ English Language & Literature
- ☐ Linguistics
- ☐ American Culture & Literature
- ☐ English Translation & Interpretation
- ☐ Other (Please indicate):

2.2 Teachers' Conceptions of Assessment Abridged Scale (TCoA-IIIA)

This **27-item survey** asks about your **beliefs** and **understanding of assessment**. Please put a tick and indicate **how much you agree/disagree** with the with each statement based on your opinion about assessment by using following scale:

- Strongly Disagree (on the left)
- Mostly Disagree
- Slightly Agree
- Moderately Agree
- Mostly Agree
- Strongly Agree (on the right)

	Strongly Disagree	Mostly Disagree	Slightly Agree	Moderately Agree	Mostly Agree	Strongly Agree
1- Assessment provides information on how well schools are doing						
2- Assessment places students into categories						
3- Assessment is a way to determine how much students have learned from teaching						
4- Assessment provides feedback to students about their performance						
5- Assessment is integrated with teaching practice						
6- Assessment results are trustworthy						
7- Assessment forces teachers to teach in a way against their beliefs						
8- Teachers conduct assessments but make little use of the results						
9- Assessment results should be treated cautiously because of measurement error						
10- Assessment is an accurate indicator of a school's quality						

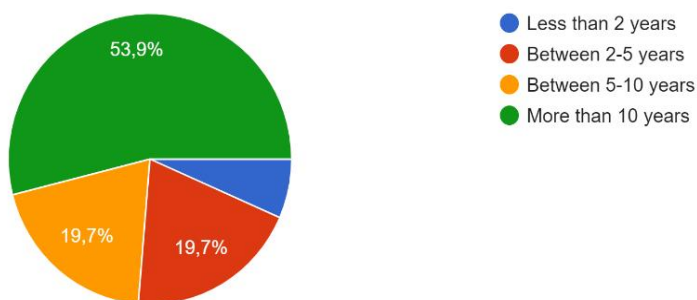
	Strongly Disagree	Mostly Disagree	Slightly Agree	Moderately Agree	Mostly Agree	Strongly Agree
11- Assessment is assigning a grade or level to student work						
12- Assessment establishes what students have learned						
13- Assessment feeds back to students their learning needs						
14- Assessment information modifies ongoing teaching of students						
15- Assessment results are consistent						
16- Assessment is unfair to students						
17- Assessment results are filed & ignored						
18- Teachers should take into account the error and imprecision in all assessment						
19- Assessment is a good way to evaluate a school						
20- Assessment determines if students meet qualifications standards						
21- Assessment measures students' higher order thinking skills						
22- Assessment helps students improve their learning						
23- Assessment allows different students to get different instruction						
24- Assessment results can be depended on						
25- Assessment interferes with teaching						
26- Assessment has little impact on teaching						
27- Assessment is an imprecise process						

Appendix 3. Responses on Google Forms

It should be noted that some forms were distributed by hand and those answers on paper were added on an Excel sheet. Below are the answers on Google Forms.

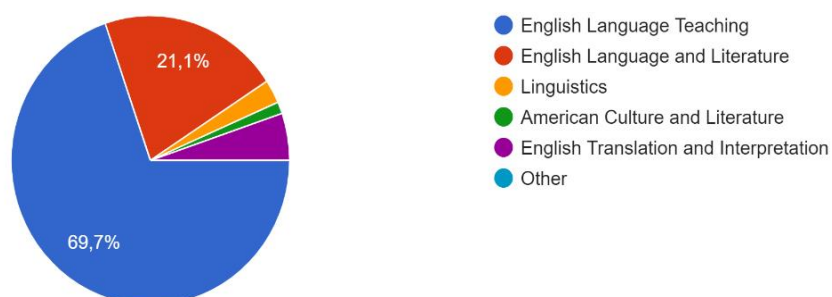
How long have you been teaching?

76 yanıt



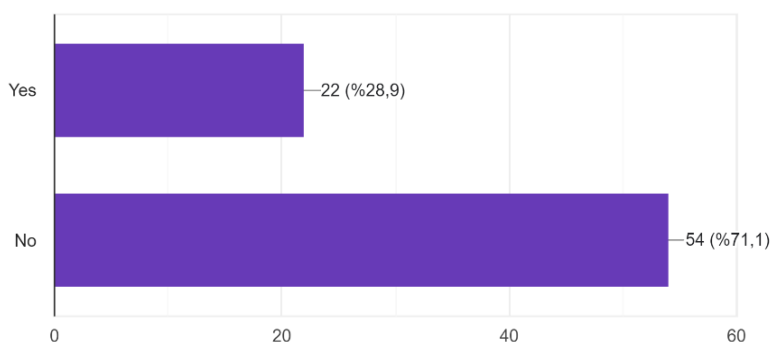
Which department did you graduate from?

76 yanıt



Would you like to take part in the focus group interview part?

76 yanıt



Distribution of All Answers to TCoA-IIIA (Google Forms & Answers on paper)

		Strongly Disagree	Mostly Disagree	Slightly Agree	Moderately Agree	Mostly Agree	Strongly Agree
1. Assessment provides information on how well schools are doing	n	1	7	12	25	44	12
	%	1.0	6.9	11.9	24.8	43.6	11.9
2. Assessment places students into categories	n	1	8	15	20	48	9
	%	1.0	7.9	14.9	19.8	47.5	8.9
3. Assessment is a way to determine how much students have learned from teaching	n	0	5	6	20	45	25
	%	0	5.0	5.9	19.8	44.6	24.8
4. Assessment provides feedback to students about their performance	n	0	6	5	14	38	38
	%	0	5.9	5.0	13.9	37.6	37.6
5. Assessment is integrated with teaching practice	n	1	2	6	17	40	35
	%	1.0	2.0	5.9	16.8	39.6	34.7
6. Assessment results are trustworthy	n	0	6	22	41	31	1
	%	0	5.9	21.8	40.6	30.7	1.0
7. Assessment forces teachers to teach in a way against their beliefs	n	8	26	27	18	16	6
	%	7.9	25.7	26.7	17.8	15.8	5.9
8. Teachers conduct assessments but make little use of the results	n	4	21	21	22	26	7
	%	4.0	20.8	20.8	21.8	25.7	6.9
9. Assessment results should be treated cautiously because of measurement error	n	3	7	10	19	31	31
	%	3.0	6.9	9.9	18.8	30.7	30.7
10. Assessment is an accurate indicator of a school's quality	n	2	15	26	46	7	5
	%	2.0	14.9	25.7	45.5	6.9	5.0
11. Assessment is assigning a grade or level to student work	n	4	8	18	33	32	6
	%	4.0	7.9	17.8	32.7	31.7	5.9
12. Assessment establishes what students have learned	n	2	5	15	32	37	10
	%	2.0	5.0	14.9	31.7	36.6	9.9
13. Assessment feeds back to students their learning needs	n	1	2	16	23	42	17
	%	1.0	2.0	15.8	22.8	41.6	16.8
14. Assessment information modifies ongoing teaching of students	n	1	3	19	26	40	12
	%	1.0	3.0	18.8	25.7	39.6	11.9
15. Assessment results are consistent	n	2	9	32	35	19	4
	%	2.0	8.9	31.7	34.7	18.8	4.0
16. Assessment is unfair to students	n	22	41	21	12	4	1
	%	21.8	40.6	20.8	11.9	4.0	1.0
17. Assessment results are filed & ignored	n	22	25	17	21	10	6
	%	21.8	24.8	16.8	20.8	9.9	5.9
18. Teachers should take into account the error and imprecision in all assessment	n	0	5	11	21	30	34
	%	0	5.0	10.9	20.8	29.7	33.7
19. Assessment is a good way to evaluate a school	n	4	16	22	37	17	5
	%	4.0	15.8	21.8	36.6	16.8	5.0
20. Assessment determines if students meet qualifications standards	n	0	5	14	33	39	10
	%	0	5.0	13.9	32.7	38.6	9.9
21. Assessment measures students' higher order thinking skills	n	4	19	22	39	11	6
	%	4.0	18.8	21.8	38.6	10.9	5.9

22. Assessment helps students improve their learning	n	0	4	13	32	37	15
	%	0	4.0	12.9	31.7	36.6	14.9
23. Assessment allows different students to get different instruction	n	8	27	19	23	17	7
	%	7.9	26.7	18.8	22.8	16.8	6.9
24. Assessment results can be depended on	n	1	8	25	41	21	5
	%	1.0	7.9	24.8	40.6	20.8	5.0
25. Assessment interferes with teaching	n	14	21	18	24	16	8
	%	13.9	20.8	17.8	23.8	15.8	7.9
26. Assessment has little impact on teaching	n	22	41	19	14	4	1
	%	21.8	40.6	18.8	13.9	4.0	1.0
27. Assessment is an imprecise process	n	17	26	25	24	5	4
	%	16.8	25.7	24.8	23.8	5.0	4.0



Appendix 4. Focus Group Invitation Form

Focus Group Interview Invitation

Dear Hocam,

I am İrem Aksu Çekiç, a master's student at Gazi University- English Language Teaching Department. Through another Google Forms link (or in pen and paper format), you completed my questionnaire for my thesis study named *"An Investigation into the Assessment Conceptions of In-Service English Language Instructors"*. I cannot thank you enough for your valuable support in that step.

I am now sending you this link because you expressed your willingness to participate in the second step of my study, which is a focus group interview. The interview is going to be conducted **online via Zoom**. As it will involve more than one person at the same time, I **kindly ask you to choose the most suitable time slot/s for you so that we can arrange a meeting.**

P.S. The focus group interview may last tentatively between 40 minutes to 1 hour including 3 people in the same group. **Your name, surname, and the recording of the interview will not be shared with third parties.** If you have any questions or concerns related to any part of the study, please do not hesitate to ask.

I appreciate your cooperation and look forward to your participation. Thank you very much for your time and effort in advance!

İlerleme durumunu kaydetmek için [Google'da oturum açın](#) Daha fazla bilgi

* Zorunlu soruyu belirtir

Please indicate your name and surname for me to contact you for the interview. *

Yanıtınız

Could you please choose the most suitable time slots for you? *

- ☐ 13.12.2023- Wednesday 18:00
- ☐ 14.12.2023- Thursday 09:00 AM
- ☐ 14.12.2023- Thursday 16:00
- ☐ 14.12.2023- Thursday 20:00
- ☐ 15.12.2023- Friday 09:00 AM
- ☐ 15.12.2023- Friday 15:00
- ☐ 16.12.2023- Saturday 15:00

Appendix 5. Sample Entries from the Researcher's Self-Reflection Journal

In the second interview, some participants mentioned that assessment is something that has to change in time. I can support this view with Coombe's article. In his/her book, there are also things concerning teacher development, and I have downloaded an article on this subject. Also, I can use Popham's article when discussing the results. **(Journal entry, February 7, 2024)**

Today while reporting my codes, I have realised that I could add another code and include it in the implications part as well, which is “meaningless practices”. I mean, Elif and Songül sometimes mention that if they change the way an application takes place, students will find it more meaningful and maybe learn more. For instance, Elif mentions departmental courses’ being irrelevant, which makes students demotivated. Songül indicates that making students choose their own presentation topics would be easier and better for them. Also, she mentions memorization many times in this regard. (Memorizing things that you do not even know the meanings of.) So, I can include those opinions under this category. Also, Leyla mentions grading can be a source of stress for students, which I will discuss in the student accountability section as well. **(Journal entry, February 22, 2024)**

I have decided to change the name of one of my codes. Actually, I have realised that teachers explain how assessment courses affected/shaped their perspective of assessment and this effect is not always negative, especially when they talk about their master’s degree courses and this makes it necessary for me to change the name of the code “inefficiency of assessment courses in shaping prospective teachers’ mindset/conceptions.” Also, when I find parts related to the washback effect, I will include them in my paper. **(Journal entry, February 29, 2024)**

Today, I have decided to do something different. I am just going to present the results in the findings section, and I will discuss them in the next chapter based on experience years of the participants and department. Also, I think it will be better to include another section by writing suggestions from instructors; maybe there is a certain pattern in their suggestions as well. **(Journal entry, March 11, 2024)**

Appendix 6. Consent Form for the Focus Group Interview

Dear participant,

I am İrem Aksu Çekiç, a master's student at Gazi University-English Language Teaching Department. You are invited to participate in the second step of my research study named "*An Investigation into the Assessment Conceptions of In-Service English Language Instructors*" after your statement of volunteering for the focus group interview in the first form provided by the researcher.

As you are familiar, this study is being conducted under the supervision of Assoc. Prof. Dr. Müzeyyen Nazlı Güngör to comprehend the assessment conceptions of in-service English language instructors as well as the assessment techniques that are used by them.

In this second phase of my study, **which is going to take place as a focus group interview**, you are going to be asked to share your opinions and beliefs on assessment and share your assessment-related experiences and practices by being involved in a group discussion and providing answers to the questions asked by the researcher.

The interview is going to be an online meeting **via Zoom consisting of 4 people** in total including the researcher. The meeting is going to be **video-recorded** and both the recording and your personal information, such as name and surname **are not going to be shared with third parties**. The information you provide is only going to be used for academic purposes **with pseudonyms**. If you have any questions/concerns related to the study and confidentiality, you are always welcome to share them with me.

You may leave the study at any time if you do not feel comfortable.

I read the information above and agree to take part in the study.

Name Surname

Signature:

Appendix 7. Focus Group Interview Questions

1. Can you please tell me about your background? (e.g. the department that you graduated from, your teaching experiences, etc.?)
2. How do you define/describe assessment?
3. To what extent is assessment important for you? Why?
4. Do you think assessment has an effect on student learning? If so, how?
5. How competent do you feel about assessing your students' language skills?
6. Did you have any assessment courses during your pre-service education? (e.g. Language Testing-Evaluation courses.)
7. What are your opinions about the assessment courses in your pre-service education years?
8. Do you think those courses affected your understanding/ beliefs of assessment? If so, how?
9. What assessment practices do you use to see your learners' progress? (e.g. portfolios, quizzes, exams, projects, peer-feedback, and so on.)
10. How do you decide on the assessment practices that you use?
11. Do you think you make use of the assessment results of students?
12. In your opinion, is there a relationship between the success of teachers or institutions and students' assessment results? How?
13. How do you think can the assessment practices be improved in your institution?
14. Do you find the current exams/procedures applied in your institution effective to determine students' proficiency levels, such as A1-B2?
15. When you consider your assessment practices from the beginning of your teaching experience until now, do you realise a change? If yes, how?

Appendix 8. Coding Process on MAXQDA

The screenshot displays the MAXQDA software interface during a coding process. The top menu bar contains various options: Giriş Sayfası, İçer aktar, Kodlar, Memolar, Değişkenler, Analiz, Karma Yöntemler, Görsel Araçlar, Raporlar, Stats, and MAXDicto. Below the menu, there are icons for creating new projects, opening existing ones, and managing code systems. The main workspace is divided into three panes. The left pane, titled 'Belgeler', shows a list of interviews (Interview 1, Interview 2, Interview 3) and clusters (Kümler). The middle pane, titled 'Kod Sistemi', displays a list of codes and their frequencies. The right pane, titled 'Belge Tarayıcısı: Interview 2 (Sayfa 8/19)', shows a text document with highlighted segments and a vertical timeline of codes. The timeline includes codes like 'My opinions on asses...', 'Inefficiency of assess...', 'School, time and curr...', 'Some steps should...', and 'Assessment is alive...'. The bottom status bar shows 'Tek Kodlama Sorgusu (VEYA kodların kombinasyonu)'.

Appendix 9. Sample Codes and Pre-determined Themes

Interview Codes	Pre-Determined Themes
Assistance for Teachers	Improvement
Including Peer Assessment and Self-Assessment	Student Accountability
Meticulous Design	Improvement
Washback Effect	School Accountability
Teacher Competencies	Irrelevance
Guide for Students	Student Accountability

Appendix 10. Ethics Committee Approval

Evrak Tarih ve Sayısı: 20.04.2023-E.638836



T.C.
GAZİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Etik Komisyonu

Sayı : E-77082166-302.08.01-638836
Konu : Bilimsel ve Eğitim Amaçlı

20.04.2023

Dağıtım Yerlerine

Üniversitemiz Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Ana Bilim Dalı **Yüksek Lisans Öğrencisi İrem AKSU'nun, Doç. Dr. Müzeyyen Nazlı GÜNGÖR'ün** danışmanlığında yürüttüğü "*An Investigation into the Assessment Conceptions of in-service English Language Instructors*" adlı tez çalışması ile ilgili konu Komisyonumuzun **04.04.2023** tarih ve **07** sayılı toplantısında görüşülmüş olup,

İlgilinin çalışmasının, yapılması planlanan yerlerden izin alınması koşuluyla yapılmasında etik açıdan bir sakınca bulunmadığına oybirliği ile karar verilmiş ve karara ilişkin imza listesi ekte gönderilmiştir.

Bilgilerinizi rica ederim.

Araştırma Kod No: 2023 - 505

Prof. Dr. İsmail KARAKAYA
Komisyon Başkanı

Ek:1 Liste

DAĞITIM

Gereği:

Sayın Doç. Dr. Müzeyyen Nazlı GÜNGÖR

Bilgi:

Eğitim Bilimleri Enstitüsü Müdürlüğüne

Belge Doğrulama Kodu :BSM7VDH772

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Takiis Adresi : <https://www.turkive.aov.tr/gazi-universitesi-ebvs>

Bu t



GAZİLİ OLMAK AYRICALIKTIR...

