

T.C.
ISTANBUL AYDIN UNIVERSITY
INSTITUTE OF GRADUATE STUDIES



**EFL INSTRUCTORS' VIEWS AND PRACTICES OF
CLASSROOM ASSESSMENT AT TURKISH UNIVERSITIES**

MASTER'S THESIS

Merve SARGIN

Department of Foreign Languages Education
English Language Teaching Program

AUGUST, 2024

T.C.
ISTANBUL AYDIN UNIVERSITY
INSTITUTE OF GRADUATE STUDIES



**EFL INSTRUCTORS' VIEWS AND PRACTICES OF
CLASSROOM ASSESSMENT AT TURKISH UNIVERSITIES**

MASTER'S THESIS

Merve SARGIN
(Y2112.021010)

Department of Foreign Languages Education
English Language Teaching Program

Thesis Advisor: Prof. Dr. Hatice SOFU

AUGUST, 2024

THESIS EXAM REPORT

Istanbul Aydın University Institute of Graduate Studies Board of Directors
22.07.2024 date and 2024/12. The thesis of EFL Instructors' Views and Practices of
Classroom Assessment at Turkish Universities, whose thesis defense exam was held
on 01/08/2024 before the jury members formed at the meeting no. Unanimity* and
Acceptance** decision was made.

JURY

1st Member (Thesis Advisor)	:Prof. Dr. Hatice Sofu
2nd Member	:Prof. Dr. Türkay Bulut
3rd Member	: Assoc. Prof. Dr. Tuba Demirkol

APPROVAL

Istanbul Aydın University Institute of Graduate Studies Board of Directors
..... date and decision no.

(*) Unanimity/Majority vote will be written in writing.

(**) Acceptance decision will be written in writing.

DECLARATION

I hereby declare with respect that the study “EFL Instructors’ Views and Practices of Classroom Assessment at Turkish Universities”, which I submitted as a Master thesis, is written without any assistance in violation of scientific ethics and traditions in all the processes from the Project phase to the conclusion of the thesis and that the works I have benefited are from those shown in the References.
(01/08/2024)

Merve SARGIN

FOREWORD

I would like to begin by expressing my deepest gratitude to my esteemed thesis advisor Prof. Hatice SOFU for her sincere support, collaboration and patience. It has been an honor to collaborate with her and learn from her. This endeavour would not have been possible without her valuable feedback and guidance.

I am also grateful to Prof. Dr. Türkay BULUT and Assist. Prof. Dr. Hülya YUMRU who enriched my learning experience with their guidance, constructive feedback and commitment to fostering a stimulating academic environment.

Moreover, I would like to thank my spouse for his never-ending support, encouragement and understanding. I am deeply grateful for his love, patience, and unwavering faith in my pursuits.

I also wish to express my appreciation to my dearest friends Feyza Özdeş whose unwavering support and friendship. She has always been a constant source of inspiration and strength.

Furthermore, I would like to extend my sincere thanks to my dearest family whose sacrifices and dedication have paved the way for my education. Your constant belief in my abilities, countless words of encouragement, and endless patience have been instrumental in my journey to completing this thesis.

August, 2024

Merve SARGIN

EFL INSTRUCTORS' VIEWS AND PRACTICES OF CLASSROOM ASSESSMENT AT TURKISH UNIVERSITIES

ABSTRACT

It has been widely acknowledged that assessment is a key component of teaching and learning and the quality of classroom assessment has impacted students' learning and the quality of teaching. Grounded in a mix method design, the present study aimed to explore the purposes, methods, and practices of classroom assessment used by EFL instructors working at the preparatory schools of Turkish universities. The quantitative data were collected via a 5-point Likert scale questionnaire which consisted of 4 sections with 57 items in total. In the purposes and methods parts, the participants stated their views using a 5-point Likert scale: a scale with five options: 1 (Never 0%), 2 (Rarely 25%), 3 (Often 50%), 4 (Usually 75%) and 5 (Always 100%). In the fourth part, they responded using a 5-point Likert scale: (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree. For collecting quantitative data, the survey method was employed. The questionnaire was administered to 100 (21 M, 79 F) EFL instructors selected by convenient sampling from different universities in Istanbul, İzmir, Bursa, Aydın, Mersin, Ankara and Eskişehir. The quantitative data obtained were analysed using SPSS (26) by descriptive statistics (frequencies). For qualitative data, semi-structured interviews were conducted in English with 10 participants who had already filled in the survey. Related to the practices and beliefs of EFL Instructors, 10 interview questions were formed after the questionnaire to clarify some answers/results. The qualitative data were analysed using content analysis. The findings indicated that the instructors use classroom assessment mainly for formative purposes: (1) to learn more about students' growth, (2) to give feedback to students during the course, (3) to diagnose the strengths and weaknesses of their students and (4) to motivate students. Also, the results revealed that while summative methods were mainly used to assess reading and writing, formative methods were used to

assess listening and speaking. The results of the survey indicated that the instructors strongly believed in AfL, in the areas of giving feedback and modifying instruction based on assessment results. However, their beliefs were not fully in line with their methods in terms of student involvement through self and peer assessment practices, and reflection on their teaching because of issues like low student proficiency, administrative restraints, tight schedules, heavy workload, negative washback effects, and time constraints.

Keywords: Classroom Assessment, Language Assessment, Assessment Purposes, Teacher Beliefs, Assessment Methods



TÜRKİYE’DE İNGİLİZCE ÖĞRETİM ELEMANLARININ SINIF İÇİ DEĞERLENDİRME ÜZERİNE GÖRÜŞ VE UYGULAMALARI

ÖZET

Değerlendirmenin öğretme ve öğrenmenin önemli bir bileşeni olduğu ve sınıf değerlendirmesinin kalitesinin öğrencilerin öğrenim ve öğretim kalitesini etkilediği yaygın olarak kabul edilmektedir. Karma yöntemler araştırma tasarımına dayanan bu çalışma, Türkiye'deki üniversitelerin hazırlık okullarında çalışan İngilizce öğretim elemanları tarafından kullanılan sınıf içi değerlendirme amaçlarını, yöntemlerini ve uygulamalarını keşfetmeyi amaçlamıştır. Nicel veriler 4 bölümden oluşan ve toplam 57 maddeden oluşan 5'li Likert ölçekli anket aracılığıyla toplanmıştır. İlk üç bölümde katılımcılar görüşlerini 5'li Likert ölçeği kullanarak belirtmiştir: 1 (Asla %0), 2 (Nadiren %25), 3 (Sıklıkla %50), 4 (Genellikle %75) ve 5 (Her zaman %100). Dördüncü bölümde ise (1) kesinlikle katılmıyorum, (2) katılmıyorum, (3) kararsızım, (4) katılıyorum ve (5) kesinlikle katılıyorum olmak üzere 5'li Likert ölçeği kullanılarak yanıt verilmiştir. Nicel verilerin toplanmasında anket yöntemi kullanılmıştır. Anket İstanbul, İzmir, Bursa, Aydın, Mersin, Ankara ve Eskişehir'deki farklı üniversitelerden kolayda örneklem yöntemi ile seçilen 100 (21 E, 79 K) İngilizce öğretim elemanına uygulanmıştır. Elde edilen niceliksel veriler SPSS (26) kullanılarak tanımlayıcı istatistikler kullanılarak analiz edilmiştir. Nitel veriler için ise anketi doldurmuş olan 10 katılımcıyla İngilizce olarak yarı yapılandırılmış görüşmeler yapılmıştır. Anketteki bazı cevapları/sonuçları netleştirmek amacıyla kullanılan ankete bağlı olarak, katılımcıların sınıf için değerlendirme uygulamaları ve inançları ile ilgili 10 görüşme sorusu oluşturulmuştur. Nitel veriler içerik analizi kullanılarak analiz edilmiştir. Bulgular, öğretmenlerin sınıf değerlendirmesini temel olarak 4 biçimlendirici amaç için kullandıklarını göstermiştir: (1) öğrencilerin gelişimi hakkında daha fazla bilgi edinmek, (2) ders sırasında öğrencilere geri bildirim vermek, (3) öğrencilerinin güçlü

ve zayıf yönlerini teşhis etmek ve (4) öğrencileri motive etmek. Ayrıca sonuçlar, genel olarak okuma ve yazma becerilerini değerlendirmek için özetleyici değerlendirme yöntemlerin kullanıldığını, dinleme ve konuşmayı değerlendirmek için ise biçimlendirici yöntemlerin kullanıldığını ortaya çıkarmıştır. Anketin sonuçları, öğretmenlerin geri bildirim verme ve değerlendirme sonuçlarına göre öğretim yöntemlerini düzenleme alanlarında, Öğrenme için Değerlendirme'ye güçlü bir şekilde inandıklarını göstermiştir. Ancak öğrencilerin düşük İngilizce seviyesi olması, idari kısıtlamalar, sıkışık ders programları, öğretim elemanlarının ağır iş yükü, sınavların olumsuz etkisi ve zaman kısıtlamaları gibi sorunlar nedeniyle, öğretim elemanlarının inançları, değerlendirme yöntemleri, öz ve akran değerlendirme uygulamaları ve kendi ders öğretimleri ile ilgili düşünme alanlarında pratiklerine yansımadağı sonucuna varılmıştır.

Anahtar Kelimeler: Sınıf Değerlendirmesi, Dil Değerlendirmesi, Değerlendirme Amaçları, Öğretmen İnançları, Değerlendirme Yöntemleri

TABLE OF CONTENTS

DECLARATION.....	i
FOREWORD.....	iii
ABSTRACT	v
ÖZET.....	vii
TABLE OF CONTENTS.....	ix
LIST OF FIGURES	xiii
LIST OF TABLES	xv
LIST OF ABBREVIATIONS	xvii
I. INTRODUCTION.....	1
A. Background of the Study.....	1
B. Statement of Problem and Purpose	3
C. Research Questions	5
D. Significance of the Study	5
E. Key Concepts and Definitions	6
II. LITERATURE REVIEW.....	7
A. Testing and Assessment in Language Teaching	7
1. Principles of Assessment	8
a. Validity.....	8
b. Reliability.....	9
c. Authenticity.....	11
d. Washback	11
e. Practicality.....	12
2. Types of Assessment.....	12
a. Summative and formative assessment	13
b. Norm-referenced and criterion-referenced assessment.....	13
c. Informal and formal assessment.....	14
d. Discrete-point and integrative assessment	15
e. Subjective and objective assessment.....	15

f. Direct and indirect assessment	16
B. The Shift from Traditional to Alternative Assessment	16
C. Classroom Assessment	19
1. Purpose of Classroom Assessment.....	21
2. Classroom Assessment Process.....	24
3. Methods of Classroom Assessment.....	30
a. Assessment of reading.....	30
b. Assessment of writing	32
c. Assessment of listening.....	34
d. Assessment of speaking	36
4. Teachers' Beliefs about Classroom Assessment Practices	40
III. METHODOLOGY	45
A. Research Design	45
B. Setting and Participants	46
C. Data Collection Instruments.....	48
1. Questionnaire about Teachers' Views, Purposes and Methods of Classroom Assessment.....	48
2. Semi-Structured Interviews.....	49
D. Data Collection Procedure	50
E. Data Analysis	51
IV. FINDINGS	53
A. Findings Regarding the Purposes of Classroom Assessment.....	53
1. Findings from the Questionnaire.....	53
2. Findings from the Interviews	55
B. Findings Regarding the Methods of Assessment	58
1. Findings from The Questionnaire	58
2. Findings from the Interviews	63
a. Methods of reading assessment.....	64
b. Methods of writing assessment	66
c. Methods of listening assessment.....	69
d. Methods of speaking assessment.....	72
C. Findings Regarding the Beliefs of EFL Instructors	75
1. Findings from the Questionnaire.....	75
2. Findings from the Interviews	79

D. Summary of the Results	88
1. Summary of the Questionnaire Results.....	88
2. Summary of the Interview Results.....	90
V. DISCUSSION AND CONCLUSION	93
A. Introduction	93
B. RQ1: For Which Purposes do EFL Instructors at Turkish Universities Use Classroom Assessment?	93
C. RQ2: Which Methods do EFL Instructors at Turkish Universities Use to Assess Four Skills (Reading, Writing, Listening, and Speaking)?	95
D. What are EFL Instructors' Beliefs about Classroom Assessment? To What Extent are the Beliefs of Instructors in Line with Their Practices?	100
E. Conclusion	104
F. Limitations of the Study.....	106
G. Suggestions for Further Research	107
VI. REFERENCES.....	109
APPENDICES	143
RESUME.....	157



LIST OF FIGURES

Figure 1. Percentages of the purposes of assessment.....	54
Figure 2. Percentages of the reading assessment methods.....	59
Figure 3. Percentages of the writing assessment methods	61
Figure 4. Percentages of the listening and speaking assessment methods.....	62
Figure 5. Types of feedback provided by instructors.....	79





LIST OF TABLES

Table 1 Old and New Paradigms	17
Table 2. Assessment Methods in Listening.....	35
Table 3. Demographic Characteristics of Questionnaire Participants	47
Table 4. Purposes of Assessment as Stated by Instructors.....	54
Table 5. Methods of Reading Assessment	58
Table 6. Methods of Writing Assessment	60
Table 7. Methods of Listening and Speaking Assessment.....	62
Table 8. Beliefs about Language Assessment.....	76
Table 9. Using Peer and Self-assessment.....	77



LIST OF ABBREVIATIONS

AaL	: Assessment as Learning
AfL	: Assessment for Learning
AoL	: Assessment of Learning
ELP	: English Language Proficiency
N	: Number of Participants
M	: Mean Value
SD	: Standard Deviation



I. INTRODUCTION

In order to set a context for this thesis, this chapter includes a background of the study, statement of problem, purpose of the study and research questions, the significance of the study and operational definitions.

A. Background of the Study

Assessment is a fundamental part of the learning process as it has a critical function in modifying instruction, diagnosing students' strengths and weaknesses related to classroom instruction and providing feedback to students to support learning. In recent years, the pedagogical theory in EFL teaching has witnessed major changes and these changes affected assessment procedures and approaches.

The shift from behaviourist teaching pedagogy to more learner-centred pedagogies paved the way for the changes in assessment. Traditional education focuses on the stimulus-response association which is based on behaviourist approach and students' learning is claimed according to the increase in the right answers in tests. On the other hand, the constructivist approach assumes that knowledge is socially constructed and emphasizes traditional assessment tools are not solely effective as they do not take the learning process into consideration (Birgin and Baki, 2007). Also, according to the constructivist approach, traditional assessment cannot provide students with a free learning environment as they restrict students' learning goals.

For many years, assessment in a classroom setting was viewed as assessment conducted inside the classroom under the guidance of the teacher (i.e., assessment external to the classroom such as school board examinations, and standardized exams) and it was assumed that the same testing procedures, activities, and methods used for large-scale testing would also be applicable to classroom assessment (Turner, 2012). However, starting with Black and William's study (1998), which highlights the critical role that assessment plays in fostering learning, the ongoing

process-based formative function of assessment has started to be taken into consideration and assessment should have a broader definition including both learning process and product (Camp, 1993) and it should give information about how to facilitate learning (Shermis and DiVesta, 2011). Thus, in classroom assessment, teachers not only act as assessors who can organize, carry out, and administer fair tests, enabling them to make educated judgments on the language proficiency of their students (Estaji and Kardoust, 2021) but also as facilitators and active agents of the assessment process (Rea-Dickins, 2004) who can use assessment information as a reference to modify their teaching and students' learning. In this respect, teachers' perceptions, knowledge and competence in classroom assessment may have a positive or impeding effect on the quality of assessment.

Recently, research on classroom assessment has gained momentum in the world. These studies mainly focus on teachers' perceptions, purposes and practices of classroom assessment as well as assessment literacy and ethics (Black and Wiliam, 1998; Brookhart and Durkin, 2003; Can, 2017; Cheng et al., 2004; Dang, 2017; Frey and Schmitt, 2010; Gönen and Akbarov, 2015; Kırkgöz et al., 2017; McMillan, 2013; Özdemir-Yılmaz and Özkan, 2017; Swai, 2019; Thomas, 2012; Yao, 2015).

Regarding the perceptions of teachers, Dang (2017) carried out research on teachers' perceptions of classroom-based assessment. To collect the data, focus group interviews were employed with teachers at the English Department of a public university in Vietnam. The results showed that the teachers perceived classroom assessment as a process that should be a part of instruction. Accordingly, they stated that an effective assessment should be objective, methodical, and criteria-based, incorporating student participation and comprehension of the assessment goals.

Related to assessment literacy, Öz and Atay (2017) investigated Turkish EFL Instructors' classroom assessment literacy. The research aimed to find out the effect of Turkish EFL Instructors' experience on their practices. The study utilized a semi-structured interview as a data collection tool. According to the results, it was found that many teachers were aware of the components of classroom assessment yet, in practice their classroom practice didn't match their perception and knowledge. Also, the results demonstrated that teachers' experience didn't have any effect on their perceptions.

Also, some researchers studied the teachers' beliefs on classroom assessment in terms of ethics. As an example, Can (2017), carried out research with 143 English teachers at Gaziantep and investigated teachers' classroom practices in terms of different variables such as academic background, experience, types of schools they work for, and the level at which they teach. The study revealed that teachers rely increasingly on paper-pencil tests as the level of education they teach rises. Teachers who work in public and private institutions also hold varying opinions regarding the ethics of the methods used in the classroom for assessment. It also showed that teachers' beliefs regarding the ethics of classroom assessment procedures are unrelated to the methods they use in their classrooms. Teachers at private schools are more likely to utilize performance assessment methods and believe them to be morally right. They also conclude that administering standardized examinations and revising exams are more morally right than educators employed by public schools. This indicates that private school instructors hold distinct moral opinions on the methods used in the classroom for assessment.

Additionally, Swai (2019) investigated the secondary school EFL teachers' purposes, methods and perceptions of classroom assessment. In order to collect data, an online questionnaire and semi-structured interviews were employed with teachers. The findings revealed that teachers utilized classroom assessment to achieve various goals, generally connected to students' learning. Secondly, for assessing reading, writing and listening, teachers preferred instructor-oriented methods more while for speaking, student-conducted methods were more frequently utilized. The data also indicate that, while instructors were favorably oriented toward formative assessment, their views may not be consistent with their reported assessment methods.

In conclusion, these studies show that, due to a variety of reasons including instructional context and ethical convictions, there is a discrepancy between instructors' perceptions and actual actions in the classroom when it comes to assessment. Although educators embrace formative techniques and understand the value of incorporating evaluation into instruction, they frequently fall short of these goals in practice.

B. Statement of Problem and Purpose

Classroom assessment is a vital aspect of successful teaching and learning

(Asamoah et al., 2023) since it comprises all of the tasks that teachers and learners perform in order to get data for the purposes of instruction and learning (Abeywickrama, 2012). Teachers spend a significant amount of their time planning and preparing assessment tools and observation methods, as well as marking, documenting, and summarizing the results in informal and formal reports in their daily teaching (Cheng et al, 2004). They constantly choose what they think is most suitable and in the best interests of the students they teach based on their expertise, their understanding of language acquisition, growth, and competency in general. According to Stiggins (2010), teachers should have a thorough grasp of classroom assessment to provide an equal emphasis to both *Assessment of* and *Assessment for* learning but it is not always easy to shift classroom assessment away from its customary summative focus (Tierney, 2006). Thus, teachers' understanding and practices might have a beneficial as well as impeding role in classroom assessment when the balance is lost (Rea- Dickens, 2004).

Although there are several studies investigating language assessment and formative assessment in the Turkish context (Burnaz, 2011; Çetin, 2011; Öztürk and Aydın, 2019; Karaman and Şahin, 2017; Hatipoğlu, 2015; Mede and Atay, 2017; Önalın and Karagül, 2018; Öz, 2014; Özkan, 2020; Gökçe, 2014 Örsdemir, 2010; Ak, 2013; Çiftçi, 2010, Özuslu, 2018), research about classroom assessment investigating the practices of EFL instructors especially in Turkey are scarce (Acar Erdol and Yıldızlı, 2018; Can, 2017; Gönen and Akbarov, 2015; Öz and Atay, 2017; Özdemir-Yılmaz and Özkan, 2017). While four of these studies are conducted at university context (Can, 2017; Gönen and Akbarov, 2015; Özdemir-Yılmaz and Özkan, 2017; Öz and Atay, 2017), one of them had a small sample with 12 instructors from one university (Öz and Atay, 2017); four of them used either quantitative or qualitative method for data collection (Can; 2017; Gönen and Akbarov, 2015; Öz and Atay; 2017; Özdemir -Yılmaz and Özkan; 2017). Moreover, while four studies focused on Turkish EFL teachers' classroom assessment practices from different aspects such as ethics or procedures, none of them investigated EFL Instructors' beliefs about classroom assessment in terms of Assessment for Learning or formative assessment. Consequently, an examination of EFL Instructors' beliefs and practices of classroom assessment is required to have a thorough understanding of the factors influencing the effectiveness of classroom

assessment.

C. Research Questions

The aim of this study was to explore the purposes, methods, beliefs and practices of classroom assessment used by EFL Instructors working at the preparatory schools of Turkish universities. Throughout the investigation, the following research questions were looked into in light of the current study's goals:

- 1- For which purposes do EFL Instructors at Turkish universities use classroom assessment?
- 2- Which methods do EFL Instructors at Turkish universities use to assess four skills (reading, writing, listening, and speaking)?
- 3- What are EFL Instructors' beliefs about Classroom Assessment? To what extent are the beliefs of instructors in line with their practices.

D. Significance of the Study

The existing body of literature on EFL instructors' views and practices of classroom assessment, especially in the Turkish context is sparse and there is a need for further research including many aspects of classroom assessment. Hence, the purpose of this study is to fill in the gap in the current literature about classroom assessment by presenting the beliefs and practices of EFL instructors at tertiary level in a Turkish context.

Furthermore, while studies on classroom assessment usually either focus on teachers' perceptions or practices of classroom assessment, this study approaches classroom assessment from many different points including purposes of classroom assessment, methods of classroom assessment for four skills and the beliefs of EFL Instructors about the formative function of classroom assessment; thus it will draw a detailed and holistic picture about classroom assessment. This way, it may (1) be a model for pre-service teachers and inform them about how to carry out effective classroom assessment by providing a detailed description of teachers' practices; (2) enhance the existing knowledge of the experienced teachers by giving them chance to compare their practices with others (3) contribute to the methodology of the literature by presenting the results of questionnaires and interviews; (4) inform the

policymakers on the factors that improve and hinder the quality of assessment.

E. Key Concepts and Definitions

1. Assessment: Assessment is (a) a set of procedures (b) designed to provide information about students' development, growth, and achievement (c) as compared with a standard (Shermis and Di Viesta, 2011).

2. Evaluation: A procedure for gathering information in order to make changes, removal, and/or approval of anything in the curriculum. (Ornstein and Hunkins, 2017).

3. Summative Assessment: Assessment which counts towards, or constitutes a final grade for, a module or course or here a pass is required for progression by the student (Bloxham and Boyd, 2007)

4. Formative Assessment: A planned process in which assessment-elicited evidence of students' status is used by Instructors to adjust their ongoing instructional procedures or by students to adjust their current learning tactics (Popham, 2009).

5. Classroom Assessment: All the acts of teachers such as gathering, analysing, and interpreting data to make decisions on learning and teaching in a classroom context (Airasian and Russell, 2008).

II. LITERATURE REVIEW

The primary aim of this study is to examine the purposes and methods of classroom assessment together with the instructors' views on classroom assessment. In this frame, this chapter will look into the difference between testing and assessment, the shift from testing to assessment and classroom assessment.

A. Testing and Assessment in Language Teaching

In educational contexts, assessment and testing are used interchangeably. Thereby, when assessment methods are concerned, they generally evoke the image of tests. However, assessment and testing are significantly different terms. According to Brown (2004), testing refers to planned administrative processes that happen at predetermined periods in a curriculum when students must use all of their resources to perform at their highest level, and when they are conscious that their responses are being recorded and assessed. Similarly, for Dikli (2003), testing is a formal, standardized means of measuring the knowledge and abilities of the students. Andhika (2011) also defines testing with its connection to scores as one of the techniques for gathering information. Besides, she emphasizes the narrow scope of testing which does not include making judgments to improve the learning based on these data.

While describing the difference between testing and assessment, Miller (2008) draw attention to the learning and decision-making process. For them, assessment refers to any actions that are carried out to gather data regarding students' learning processes and to make essential judgments based on these data. According to Ezir (2013), testing is a one-time, formal, and frequently standardized way of measuring students' performance, so everyone takes the test in the same circumstances and the test items, administration, and scoring methods are the same for everyone, whereas assessment is a broader term that entails a variety of formal and informal means of collecting information in various times and contexts. When viewed from this respect, assessment refers to collecting data from multiple sources

and examining it to have a better idea of what students know, comprehend, and can do with their knowledge as a consequence of their educational experiences. (Huba and Fee, 2000).

1. Principles of Assessment

According to Brown (2004), there are five main principles of language assessment: validity, reliability, authenticity, washback, and practicality. They will be discussed in more detail in the following parts.

a. Validity

One of the most crucial components of assessment that affects the quality of assessment is validity. Roberto Lado (1961), who first applied the term validity to language assessment in 1961, encapsulates the concept in a statement: if a test measures what it is supposed to measure, it is valid. To illustrate, a listening test that consists of true/false items cannot be regarded as valid for assessing writing skills because they are different skills. Additionally, in a general sense, validity deals with the issue of what test results measure and what they predict (Guilford, 1954). To be more specific, if a test is valid, someone who looks at the test score can predict the participant's proficiency in the subject matter the test has covered. Thus, the validity provides the testers with consistent measurement and reasonable judgment (Siddiek, 2010).

Four types of evidence that are commonly used for validity are face validity, content validity, construct validity, and criterion-related validity (Lammers and Badia, 2005). Face validity is the extent to which a test seems to be appropriate in the eyes of the test-takers. In other words, a test has face validity if the content only seems relevant to the test-taker in terms of its physical features such as readability, uniformity of style and formatting, and the use of clear language etc (Taherdoost, 2016).

Content validity is about the level of correspondence between the content of the test and the desired construct. Namely, to demonstrate that a state's European Language Proficiency (ELP) test accurately assesses students' English language proficiency, one can look at how closely the test's content corresponds to the concept of ELP as it is described in the state's ELP standards (Wolf et al, 2008). To give valid evidence to support the goal of assessing the intended constructs of the test, test

creators commonly perform an expert evaluation of the content of the items (Fulcher and Davidson, 2007).

Construct validity is about how well a test score can be interpreted as a predictor of the skill or construct that educators or test makers wish to measure (Bachman and Palmer, 1996). A test's overall validity greatly depends on construct validity (Coombe, 2018) since evidence of the test's construct validity contains information including its content, how test results compare to other factors, as well as its internal organization and reliability (Sharon, 2021). A speaking ability (or the construct), for example, involves several features (or subconstructs), including grammar, accent, fluency, and comprehension, to establish the construct validity of an oral exam (Li, 2011).

Criterion-related validity is a type of evidence whereby newly developed test findings are compared with those of previously developed tests of the same skill or topic to determine the degree of correlation between test takers' scores. If a test accurately assesses a specific element or skill, it ought to be highly correlated with some other tests that assess the same element or skill (Coombe et al, 2012). There are two types of criterion-related validity: concurrent validity and predictive validity. For concurrent validity evidence, the test and criterion data are administered at the same time and the scores are compared whereas, for predictive validity, a time interval is required between the administration of the test and the scores (Fraenkel et al., 2012).

b. Reliability

Reliability is a key component of standardized testing, and when the test is reliable, the participants obtain nearly the same score after they take the test many times (Noda, 2003). Thus, the term reliability describes how consistent the test results are after being administered at different times and conditions (Harmer, 2007).

There are four methods to obtain reliability. These are;

- *Test-Retest Method* - This method means giving the same test to the same group twice after a certain amount of time (ranging from several minutes to several years) has passed.
- *Equivalent-Form Method* - The equivalent-forms approach involves administering two distinct but equivalent (also known as alternate or parallel) forms of an instrument to the same group of people at the same

time. Despite differences in question sets, they sample the same material in terms of difficulty and content are prepared independently of one another. (Miller et al., 2008).

- *Methods to Estimate Internal Consistency*

- *Split-Half Method* – This method is used to measure the internal consistency of the test. After the test takers take the test and get the results, the test items are divided into two groups as odd number and even number items, and are scored separately. The correlation of the two parts gives the degree of reliability. (Fraenkel et al., 2012; Miller et al., 2008).

- *Methods Based on Item Variances: Kuder- Richardson Reliability Coefficients (KR-20) and Cronbach Alpha.* (Kaçar, 2023). The coefficients KR-20 and alpha are the same; but, when used for dichotomously scored items, it calls KR-20, and when used for polytomous scored items, it calls alpha (Tan, 2009).

The reliability of the test is also affected by numerous factors including the test taker, the rater, the test itself, and the conditions where the test is administered. One of the major factors that affect reliability is *student-related reliability* - psychological and physical conditions of the test taker such as tiredness, lack of motivation, and sickness (Coombe et al., 2007; Heaton, 1988).

In addition to this, *the rater reliability* has an impact on the reliability of the test. Rater reliability can be applied to both inter and intra-raters. While inter-rater reliability refers to the consistency of the scores of at least two different raters for the same test, intra-rater reliability is the level of agreement between many diagnostic test administrations carried out by a single rater in different time intervals (Kayapınar, 2014). Among the possible factors that hinder the rater from assessing the examiner reliably are technical factors, which are related to the scoring criteria, the experience of the scorer, or personal factors like subjectivity and bias (Rashid and Mahmood, 2020).

Thirdly, *test reliability* is about the time and duration of the test as well as the test items and instructions. The duration of the tests should not be too long or too

short so as not to cause boredom or exhaustion. Furthermore, unambiguous items or unclear instructions may confuse the test takers and this may prevent them from demonstrating their language ability (Hughes, 1989).

Test administration reliability relates to the physical circumstances under which the test is given. Variations in copying, the amount of noise, temperature, levels of light, and the state of the seating are a few examples that affect test administration reliability negatively (Brown, 2004).

c. Authenticity

Another crucial factor that must be considered in the test design is authenticity. In order to measure language learners' cognitive processes together with their linguistic abilities and anticipate how well they will succeed in language usage in the future, it is necessary to establish authenticity (Hasrol et al., 2022). Creating assessment activities that closely resemble language use in everyday situations is the key to being authentic in language assessment. This necessitates developing assignments and resources that are relatable, significant, and suitable for test-takers' cultural backgrounds. To achieve this aim, teachers need to stimulate genuine communication scenarios that lead to dynamic interactions. Furthermore, in the exams, it is better to use a variety of language skills together (e.g. reading and writing integrated writing assessment tasks). Also, dynamic interactions should be permitted in authentic examinations. (Carroll, 1980)

A few examples of authentic assessment that teachers can use are oral interviews, in which they question students about their interests and activities; storytelling in which students relate their experiences; projects in which students can select the topics themselves and portfolios; which are concentrated collections of student work that demonstrate students' growth over time (O'Malley and Pierce, 1996).

d. Washback

Generally, tests are thought to be significant predictors of the students, classroom behaviour, procedures, and policies within the school, educational system, or society at large (Alderson and Wall, 1993). The effects of tests on instruction and learning are called "washback" or "backwash". Bachman and Palmer (1996) define the effects of washback at the macro and micro levels. At the micro level, washback

describes how much a test impacts what happens in the classroom, primarily in terms of curriculum innovation and changes, teacher approaches, and students' learning. On the other hand, at a macro level, washback is associated with how much an exam affects society as a whole, including the government, policy-making, education, publishing, general opportunities, and parental expectations of their children.

Washback is typically assessed to determine whether it has a beneficial or detrimental effect on the degree to which it promotes or inhibits forms of teaching or learning (Green, 2013). When teachers concentrate all of their lesson plans on the subjects and skills that will be assessed on a given test, they might overlook other crucial facets of learning and narrow the curriculum (Cooley, 1991). Similarly, when there is a potential mismatch between the stated teaching objectives and the focus of assessment, it may result in abandoning classroom objectives in favour of test preparation (Mckinley and Thompson, 2018).

e. Practicality

The final test quality that must be considered during assessment is practicality. Practicality refers to the relation between the necessary resources and the available resources that are needed during designing, developing, and applying tests. In other words, the test is practical if the resource requirements of the test specifications do not surpass the resources that are available at any point throughout the test's development. To be more specific, the resources may include time, administrative logistics, materials, and personnel. (Bachman and Palmer, 1996). In this respect, good tests need to be feasible. The tests are not easy to administer when they require specialized settings or pricey equipment. Similarly, subjective tests like writing or speaking are difficult to score because they require several raters and raise questions about their impartiality. However, it is not recommended to choose a test's purpose only based on practicality. Namely, even though we have recruited many raters, a multiple-choice test is not supposed to be used for composition rating because it is simple to grade. (Şengül, 2022).

2. Types of Assessment

Teachers combine different types of assessments to evaluate students' learning during classroom assessment.

a. Summative and formative assessment

Summative assessment is the assessment of a student's academic achievement at the end of a specific period of instruction (e.g. a unit, lesson, school term or an academic year, etc.) in the form of tests including multiple-choice, open-ended, true/false, short-answer questions, essays or high or low-stake assessments (Mulvenon, 2009; McMillan, 2013).

The priority of formative assessment is the ongoing development of students' learning (Brown, 2004). With this aim, students and teachers act as partners. Therefore, the role of teachers in formative assessment is supporting the learning process by integrating teaching with the desired outcome; determining the specific learning requirements of learners or groups of learners; choosing and modifying resources and materials; differentiating learning to help each learner progress at their own learning and giving immediate feedback and guidance to learners (Earl, and Katz, 2006).

The process of formative assessment starts with deciding on learning outcomes and success criteria. Before the instruction, students are made aware of the criteria together with strong and weak samples so that they know what they need to achieve. Since the assessment is embedded in instruction, teachers have the opportunity to collect evidence via any kind of classroom activity (quizzes, classroom observations, oral question-answers, conferencing etc.) to keep track of the students. The correct interpretation of the evidence helps close the gap between where the students are and where they desire to be. (Heritage, 2010). Feedback is also an important component of formative assessment. While in summative assessment feedback is provided only through grades and teachers do not give students many opportunities to revise and improve their work (Fisher and Frey, 2013), in formative assessment, students have a chance to give feedback, respond to the feedback they receive and put it into practice in different stages of learning (Carless and Winstone, 2019).

b. Norm-referenced and criterion-referenced assessment

In norm-referenced assessment, a student's performance is evaluated by comparing his / her score to the other students' scores who have taken the same test (Crocker and Algina, 2008). This kind of assessment does not directly give

information about what students can do in language or what they achieved as it is generally used to determine students' proficiency by ranking them according to others' performance (Hughes, 1989).

In contrast, in criterion-referenced assessments, students' performances are assessed according to a predetermined standard or criterion (Richard and Schmidt 2002). This type of assessment provides comprehensive data on a student's performance on each learning objective, in alignment with the objectives of the curriculum or language program (Mirmakhmudova, 2021). Furthermore, in criterion-referenced assessment, students are given a set of performance objectives to work toward at their own pace. This keeps students motivated and encourages them to work toward their "personal best" as opposed to competing with other students (McNamara, 2000). Criteria-referenced assessments are also useful for identifying learning difficulties, assessing program quality in education and social action and tracking student development in goal-based learning environments. Nevertheless, criterion-referenced assessment may have some pitfalls. Firstly, it may be challenging to write criteria to assess some skills that are not easy to measure. Secondly, as the focus is the performance, it does not give room for transferring knowledge (Hambleton et al., 1978).

c. Informal and formal assessment

Formal assessment is a method of planned assessment by systematically collecting data to determine and document students' knowledge and abilities via quizzes, tests or presentations (Brown, 2004; Nurhayati et al., 2020) and it serves the purpose of measuring students' success and reporting student progress to parents, school administrators, and the students themselves (DoE, 2007).

Unlike formal assessment, informal assessment is a less structured, not graded, and unplanned form of assessment. Informal assessment encompasses a variety of unplanned practices such as checking understanding, giving oral or written feedback, making encouraging comments on students' performance, etc. Teachers informally monitor their students in the classroom every day and collect information for many different purposes such as to track student development, to guide instruction or to have detailed information about students learning (Brown, 2004; Spinelli, 2000).

d. Discrete-point and integrative assessment

The fundamental principle of the discrete-point assessment is assessing each language point independently, whether it be pronunciation, grammar, vocabulary, or other linguistic features. Advocates of this technique believe that language is a system made up of an endless number of items and that an accurate assessment of the learner's language competency can be obtained by assessing a representative sample of these items like phonology, morphology, lexicon, and syntax (Farhadi, 1979). In this type of assessment, the focus is linguistic accuracy rather than comprehension (Schulz, 1977) and it can be beneficial when used to test particular elements of language during classroom instruction (Carr, 2011).

While discrete-point assessment aims to test language knowledge one bit at a time, integrative assessment aims to assess a learner's ability to utilize several pieces of language at once. Thus, while discrete things break down language competence, integrative tests put it back together. In integrative assessment, every language skill is equally important and it is necessary to assess skills in a more holistic and authentic manner in a context rather than isolated units because in order to communicate, it is necessary to use more than one skill (Adnan et al., 2019). In this respect, even to assess grammar, several components need to be used in the process such as vocabulary, grammar, discourse, reading and writing (Hidri, 2018).

e. Subjective and objective assessment

Subjective and objective assessments are more related to the type of scoring used in the assessment. Subjective assessments or free-answer tests such as essays, presentations, and portfolios are usually used in assessing productive skills and hence closer to real-life tasks. They require test-takers to use their creativity. Nonetheless, they are problematic in terms of rater reliability as they are open to interpretation (Huddleston, 1954). Therefore, to eliminate any possibility of subjectivity, the tests should be graded by two separate raters who are not aware of each other's scoring schemes (Karim and Haq, 2014).

Objective tests consist of multiple-choice, true/false, matching, and fill-in-the-blank items etc. so the scoring process does not leave any room for the test-taker's opinion. The items often offer a single right answer and require the respondent to pick the proper response from a list of possibilities so they are more

reliable. Furthermore, thanks to the pre-prepared key, the raters can score more easily and consistently (Shaban, 2014).

f. Direct and indirect assessment

Direct assessment or testing is the assessment of students' skills that are intended to be assessed based on their products or performances (Fraidan and Mulhim, 2012). In this respect, it is more applicable to the assessment of speaking and writing skills since they are productive skills and it is easier to establish circumstances that generate desired behaviours we can observe and assess (Hughes, 2003). Indirect assessment relies on objective items such as multiple-choice or true-false items to assess the underlying skills (Qian, 2019). To illustrate, a writing test that tests students' knowledge of mechanical aspects of writing such as grammar, organization, and syntax without requiring students to write is a way of assessing students' writing skills indirectly (Heck and Crislip, 2001).

B. The Shift from Traditional to Alternative Assessment

The term traditional assessment is a general term referring to traditional testing methods in written forms such as standardized and classroom achievement exams or quizzes which usually contain true/false, multiple choice, fill-in-the-blank, short answer questions or essays (Belle, 1999; Quansah, 2018).

Using traditional assessment techniques in education has both benefits and limitations. Traditional assessment test items are generally simple to administer, score, and analyse. They can be given out to a lot of students at once, and evaluated rather easily—in certain situations, even by computers. In other words, they are time-saving and economical. Finally, they are objective and their reliability is typically quite high (Tanner, 2001).

Besides advantages, traditional assessment has also some limitations. First of all, as they are product-oriented, one-shot and norm-referenced, they fail to consider the progress of students. Furthermore, the traditional assessment test items are limited while testing students' higher-order or productive skills. Last but not least, they are indirect and inauthentic (Bailey, 1998; Franklin, 2002).

As a result of changes in teaching techniques, methodologies, and views, language assessment practices have evolved over time and thus assessment

procedures have shifted from traditional to more communicative and student-centred ones. The table below represents the shifting paradigm in language teaching by comparing the traditional and new teaching foreign language paradigms (Tosuncuoğlu, 2018, p: 164).

Table 1 Old and New Paradigms

Old teaching foreign language paradigm	New teaching foreign language paradigm
Focus on language	Focus on communication
Isolated skills	Integrated skills
Teacher-centred	Learner-centered
Emphasis on output	Emphasis on procedure
Tests only test	Tests to teach

Since teachers employ a range of formal and informal techniques of assessment in the classroom for placement, diagnosis, formative, and summative purposes, traditional assessment is now recognized as only one type of assessment. Nowadays, a lot of academics refuse to concede that there is only one way to collect data on student learning. (Kulieke et al., 1990). The realization that assessment is not just a technical, psychometric field of knowledge has also coincided with this transformation and assessment is more clearly seen as a social practice with unique, intricate political and ethical implications, like all other parts of education (O'Loughlin, 2006).

Constructivist theory paved the way for creating more student-centred learning environments and enabled teachers to realize that assessment processes should incorporate the learning process too (Yurdabakan, 2011). Constructivism is based on three basic principles: (1) learners construct knowledge based on their experiences and prior knowledge (Alanazi, 2016); (2) learners are active participants in the learning process (Amineh and Asl, 2015); (3) learners build knowledge by socially interacting with others (Pritchard and Woollard, 2010). Based on these premises, Paily (2013) emphasizes that in a constructivist classroom, teachers are responsible for promoting and scaffolding learning by focusing on the learning process, creating opportunities for exploration, fostering collaboration and teamwork, taking students' experiences and multiple points of views into consideration,

providing an authentic learning environment and encouraging students to reflect on their learning.

According to Ahmad and Jamil (2020), the constructivist approach has revealed the inadequacy of using solely traditional product-oriented testing methods to create opportunities for learning as they do not take the learning process and students' active role in the construction of knowledge into account. Supporting this view, Huearta-Macias (1995) states that the traditional way of assessment fails to include the daily classroom activities in the assessment process; they do not provide feedback to every student on his/her strengths and weaknesses and they lack alternatives to evaluate student progress. In addition, Gömleksiz and Erkan (2016) point out that traditional assessment methods have led to some issues with teaching, such as the following: (1) assessment tools fail to measure students' accomplishments concerning the real world and they are used in a short period, which makes it difficult for students to solve problems in the real world. (2) they are insufficient to demonstrate students' abilities (3) they prevent students from accessing resources and conducting research; (4) the findings of test results are primarily used for marking rather than giving feedback to students (Erdol, 2018)

Although there has been a movement from traditional assessment toward alternative assessment, this does not mean that traditional assessment methods and tools have been abandoned at all. There is still recent research supporting or showing it to be a favoured option for several reasons. For instance, Marissa Prongsirikul (2018) conducted research on teachers' and students' perceptions of traditional and alternative assessment types at the tertiary level. The results demonstrated that traditional evaluation methods, particularly in terms of their validity and reliability, are often valued more highly by teachers and students. Another study by Önalın and Gürsoy (2020) investigated Turkish private school language teachers' perceptions of assessment and the findings revealed that teachers feel more confident while using traditional assessment in their classes. Ulker (2021) aimed to find out the frequency of usage of authentic and traditional assessments. The findings showed that although the teachers have positive perceptions of authentic assessment, in practice they rely more on traditional methods. According to Dikli (2003), many teachers favor using tests since they are easier and less time-consuming to apply.

C. Classroom Assessment

There is no acknowledged standard definition of classroom assessment, yet it generally refers to all the acts of teachers such as gathering, analysing, and interpreting data to make decisions on learning and teaching in a classroom context (Airasian and Russell, 2008). Classroom assessment is based on Assessment of Learning (AoL), Assessment *for* Learning (AfL) and Assessment *as* Learning (AaL).

Assessment of Learning (AoL) refers to the traditional assessment methods of summative assessment which aim to report achievement typically by defining where the student stands concerning a set of criteria, by measuring the level of understanding of a student on specific content or by indicating learners' relative standing in relation to others (Earl, 2003; Dunn and Mulvenon, 2009). Rather than focusing on the learning process, AoL is thought to primarily focus on testing and reporting students' academic achievement to various stakeholders such as teachers and administrators, families, and students (Lin and Lin, 2019).

While AoL allows the assessor to look back and see how well the student has succeeded at meeting the objectives, it does not indicate how to proceed in the future (Brown, 2004). This kind of feedback only informs if the students pass or fail, but to foster learning, students need to get feedback and be informed on how they to close the gap between their current and desired states (Heritage, 2010). Assessment *for* learning closes this gap.

Assessment *for* learning is the process of tracking learning while the assessment is occurring in everyday classroom practices via using formal and informal ways of assessment through identifying, analysing and reflecting on the systematic information obtained from learners (dialogue, the way of responding to tasks and questions,) and from the teachers (observation) to guide instructional decisions (Klenowski, 2009). By looking for evidence of learning and interpreting it, teachers determine where learners are in their learning, where they need to go, and what is the best way to get there (Assessment Reform Group, 2002).

When compared with AoL which concentrates just on how to measure student achievement, AfL combines teaching, learning, and assessment and highlights the vital role that assessment plays in promoting and increasing students' autonomous learning. The adoption of AfL starts with understanding the relationship between

teaching and learning, specifically by knowing how to utilize assessment to guide teaching and learning (Chen, and Bonner, 2020). The emphasis on contextual, classroom interaction rather than individual learner characteristics differentiates AfL from AoL. This is a major distinction; it focuses on what is being learned as well as the quality of classroom interactions. As the learning theory approach supporting AfL is rooted in "social constructivism", which aims to maintain a balance between the individual and cultural aspects of learning, learners are the active participants of the process and assessment serves a formative function (Stobart, 2008).

The idea of Assessment *for* learning is linked to the presumption that both teachers and students need to understand the direction in which learning is going. AfL thus serves as a tool for determining the next action to be taken (Hargreaves, 2005). Criterion-reference assessment and descriptive feedback are therefore the key elements of this process. Rather than comparing students' scores with each other, teachers negotiate learning outcomes and success criteria and provide samples of assessment to see if the criteria are met or if the learning outcomes are fulfilled. To achieve this, learners need to understand the standards that are aimed for (Stiggins and Chappuis, 2005). While instruction is taking place, teachers constantly search for evidence of how student learning is progressing toward a goal.

In AfL, feedback functions more as a continuous commentary than a measurement. It allows students to examine their performance and consider if they are on the right path without labelling or restricting their performance. Thus, the emphasis changes from "measuring" as an ultimate purpose to "assessment" as a continuous process (Wiggins, 1993). Teachers provide learners descriptive feedback on how their learning is progressing concerning the success criteria and guide students on how to move forward and close the gap (Heritage, 2010). Besides, AfL, teachers are not the only feedback providers. They are more like "the 'partners' of learning rather than the 'drivers' of what to do" (Harlan and Winter, 2004: 406).

Both students and instructors benefit from AfL because it gives them clear information about how to enhance their performance (Stiggins, 2007). While monitoring their instruction, if teachers see the learning is not progressing, they can adjust their instruction in cases where students need scaffolding (Popham, 2014). Similarly, AfL helps students evaluate and monitor their own learning and be more motivated. Through self-assessment and peer-assessment they assess themselves and

each other. This helps them understand where they are, what they achieved well and what they need to work on to do better. As students gain experience, they set goals for the next step in their learning (Stiggins and Chappuis, 2005).

Assessment *as* learning is a form of AfL in terms of its emphasis on using the assessment for facilitating learning. While in AfL, the teacher acts as a guide to enhance learning and teaching, AaL is a type of assessment that is aimed at encouraging learners' continual development of cognitive and metacognitive skills in self-evaluation (Lam, 2015).

AaL is based on the idea that learning is more than merely passing ideas from one person to another; it is an active process of cognitive reshaping. To be actively engaged in constructing their own understanding and being independent learners, learners must know how to be critical assessors who make sense of material, link it to past knowledge, apply it for new learning, and assess themselves and others through self and peer assessments on an ongoing basis (Earl and Katz 2006). In addition to that, learners must use their metacognitive skills to monitor and reflect on their own learning by deciding what to do next, monitoring the consequences, and reviewing and altering their way of thinking. (Brown, 1985; Taras, 2010). In this respect, in AaL, students engage in the process of observing themselves, reflecting on their learning and evaluating their progress. Thus, the emphasis of AaL is not only on how much a student's learning has advanced but also on to what extent his/her skills have developed (Stiggins, 2006).

In the following sections, classroom assessment will be discussed in the frame of the purpose of assessment, classroom assessment process, methods of assessment, and teachers' beliefs about classroom assessment practices.

1. Purpose of Classroom Assessment

Assessment serves a variety of multifaceted and connected purposes. Therefore, many studies are focusing on different purposes. Muñoz (2012) in his study, focused on pedagogical and administrative purposes. In the study, it is stated that the pedagogical purpose of assessment is achieved through formative assessment which helps teachers make decisions regarding their teaching and students' learning. Likewise, Quilter and Gallini (2000) pointed out teachers use both formative and summative assessment methods in their classrooms. The goal of administrative

assessment on the other hand is summative purposes since the aim is to report the outcomes of instruction to various stakeholders including teachers, administrators, and institutions.

In their study, Breen (1997) looked into the purposes, methods, and practices of assessment used by ESL/EFL instructors. According to the results, the ESL/EFL instructors use assessment to;

- *Keep track of progress and guide instructional strategies:* teachers use many techniques such as anecdotal notes, group discussions, and general observation. They employ criteria and assignments, watch certain classroom activities, and have conversations with ESL students. Systematic documentation with the use of tasks and checklists is frequently used to supplement informal observations. To help students comprehend their development and help them establish objectives, teachers frequently involve students in the assessment process. Teachers may use this data to create lessons, assign students to groups, and determine which students require more assistance.
- *To share students' progress with their parents and other teachers:* parents get reports from instructors in a variety of formats, from simple checklists with checkboxes to lengthy descriptions detailing a student's progress across the curriculum. Different teachers use work samples for assessment; some make notes on finished assignments, some evaluate them using pre-set standards, while yet others emphasize the value of autonomous completion. The reports fulfil many functions, including updating parents on their child's development, promoting parental participation, and issuing statements to the public on school learning.
- *To inform and make judgments about students' future support:* to ascertain if a student is ready for mainstream education or needs further help, a variety of assessment approaches are utilized, including band scales and diagnostic testing. Depending on a student's perceived success level, teachers make difficult decisions about when to stop supporting them or not. Some schools use joint assessment between ESL experts and other instructors to establish roles and degrees of support.

Informal decision-making frequently enhances formal processes. Educators might utilize evaluations to ask for specialized assistance, such as speech therapy.

- *To inform the school administration:* in various institutions, assessment is used for summative purposes by the principals and they collect students' reports including a numerical grade at the end of instruction.
- *To inform educational systems at a state level:*

Lambert and Lines (2000) have a similar approach to the purpose of assessment. For them, assessment serves the following purposes: to give teachers and students feedback on their progress so that future learning can be supported; to provide information about students' accomplishments at various points during and after school; to enable qualification-based selection and raise standards; and to contribute to the body of data used to make decisions about the effectiveness or quality of individuals and institutions within the system as a whole.

For Cheng and Fox (2017) assessment has mainly three purposes:

- *Instructional:* Teachers use assessment to collect data about students' learning, understanding, and skills to prepare and modify their lesson plans based on students' current learning.
- *Student-centred:* Teachers use assessment for diagnostic or placement purposes. This assessment aims to determine individual students' weaknesses and strengths and to make decisions about instruction by using this information to group them according to factors like similar skill development, learning styles, or proficiency levels (grouping). Teachers modify course instruction using diagnostic assessments to better match each student's individual learning requirements.
- *Administrative:* Assessment refers to a high-stakes, summative assessment by using numerical or ordinal scales to certify students' accomplishments at the end of instruction. This information is used to publicly demonstrate what a student has accomplished, which affects motivation, self-perception, parental expectations, curriculum, and social interaction. From this respect, assessment is examined in light of its wider consequences on several facets of

education and the student experience.

To conclude, classroom assessment is utilized for a variety of purposes, including grading evidence, monitoring student progress, parent reporting, teaching improvement, remediation, identifying student misunderstandings, student motivation, and teacher and educational accountability (McMillan, 2013). These purposes can be broadly grouped as formative and summative. Despite these two assessment purposes seeming distinct from each other, the distinction between SA and FA is not always as straightforward as it is supposed to be since the same assessment data might be utilized for various reasons at various periods. However, the primary goal of all assessments even should facilitate learning in a culture that is centred on assessment for learning. (Yan et al., 2023).

2. Classroom Assessment Process

Classroom assessment requires several activities such as selecting proper methods for assessing students, defining performance criteria, conducting and grading exams, making interpretations of results, announcing the results, planning the instruments, and evaluating the results to make decisions related to learning and teaching (Zhang and Burry-Stock, 2003).

The classroom assessment process starts with careful planning. In this stage, teachers think about what they plan to teach and how to assess the performance of students. Thoughtful planning takes into account elements such as the purpose of assessment, types of assessment methods that match with the purpose, strategies to collect evidence and the effect of results on the instruction and the frequency of assessment (Shermis and Divesta, 2011).

Teachers first select the learning goal(s) for a lesson or a series of sessions and establish the success criteria before beginning the assessment process. Learning goals or intentions are documented descriptions of the expected skills and knowledge during the course of lessons (Adam, 2004). Learning goals have many benefits for teachers and students. By using them, teachers are able to select proper content and strategies, explain the purpose of a given activity or course and plan the appropriate assessment methods (Harden, 2002). Learning goals also support students in staying focused on what they are learning and improve their understanding by being familiar with what they will learn (Kumpas-Lenk et al., 2018; Millar and Hames, 2006).

In schools where summative assessment is dominant, teachers experience a negative washback effect. Therefore, focus on performance goals, which are met by doing well on exams rather than learning goals which encourage students to focus on the activity and comprehending or mastering skills (OECD, 2008). On the other hand, learning goals make students learn more efficiently by letting them know where they stand and making the content more accessible to them. They also make what students expect to learn or achieve clear. For students to understand what is expected from them, learning goals need to be clear, brief and manageable. For this reason, teachers should avoid setting too general and complex learning goals (Crichton and McDaid, 2016; Lindstrom and Drolet, 2017)

Success criteria are closely connected with learning goals. It defines what students need to know and be able to accomplish in order to show that they have met the learning goals for the day. Likewise, they make the learning goal apparent to both teachers and students (Almarode et al., 2021). For students to apply success criteria effectively, teachers need to include small steps, make success criteria visible to students and include students in the process by discussing the criteria with them rather than telling the statements and expecting students to understand them (Willis et al., 2003). Hence, it is equally important to use simple language, and monitor students well so that they can clarify and explain any problematic points before assessment as misunderstandings may cause students to fail in achieving learning goals (Vickerman, 2009). Another important point related to success criteria is providing weak and strong models to students. By closely examining and contrasting models according to the level, using marking schemes, and creating their own success criteria, students learn about quality and assessment criteria better (Swaffield, 2011).

During the course of instruction, teachers collect valid and reliable evidence to know if the students' learning is progressing toward the learning goals and how they need to respond to the learning to improve it (Bell and Cowie, 2000). Eliciting is the process by which a teacher obtains data regarding the learning of their students, whether it be incidental (unplanned), purposeful (planned), or a mix of the two (Pai, 2018).

Purposive evidence refers to information that is intentionally elicited by a teacher or other designated person in order to demonstrate a student's proficiency in

a certain area. It includes a direct way of asking questions in oral or written forms. The disadvantage of these methods is that they do not guarantee that students demonstrate their true knowledge. Incidental evidence on the other hand is based on a teacher's observations of students on day-to-day activities. A teacher incidentally may notice a student has some skills or knowledge without planning to learn it (Black and William, 1998). According to Heritage (2010), plans for gathering evidence should be made before starting teaching since it will allow teachers to modify their lessons based on the students' answers. Even in the case of incidental evidence, teachers need to think about the necessity of adjustment of teaching.

Teachers can collect evidence according to the purpose of using them. When teachers collect evidence for summative purposes, they make exams, observe students or collect portfolios. For formative assessment, teachers observe students' behaviours or performances within the classroom, examine the product of their work (written or oral), embed assignments that call for specific knowledge and skills in routine curriculum and administer informal tests and quizzes (Dollin et al, 2018). As discussed in the section on purposes of assessment, evidence for formative assessment is used to improve teaching and learning. While summative evidence gives information about what students have achieved to the present date, formative purpose implies future learning by using evidence to make decisions to take the next step (Heritage, 2013). William (2010) emphasizes that evidence does not only provide information about where the student is and where s/he should be, or there is a gap between these two, but also it needs to specify the kinds of actions that need to be taken by the teacher and students to reach the desired performance.

After gathering evidence, both teachers and students interpret evidence based on norm-referenced, criterion-referenced standards (Mavrommatis, 1997). Accurate interpretation of evidence depends on the understanding of the skills and challenges that students face, as well as an analytical and procedural understanding of assessment and evaluating procedures. Teachers need to examine and synthesize various forms of student learning evidence correctly in order to modify teaching or offer feedback to the students to identify and close the gap (Brookhart, 2008; Sadler, 1989).

Feedback is a scaffolding mechanism that promotes an ongoing process of student learning; without feedback, classroom assessment cannot be conducted

effectively and students' opportunities of enhancing their learning are greatly diminished (García et al., 2015). Therefore, it has a vital role in classroom assessment (McFadzien, 2015).

There is a plethora of research on the positive effect of feedback on students' motivation and learning (Acarol, 2024); Hamidun et al., 2012; Kunnan and Jang, 2009; Zainuddin; 2004) and improving teaching (Black and William, 2009). Feedback can be given in a variety of ways, including written, oral, checklist, conferencing with students, total test score and a letter grade; by different agents (teachers, students themselves, peers) and for different purposes (formative and summative).

Generally, in classroom assessment, feedback is used for both summative and formative purposes. Feedback for summative purposes is used to inform students about the total score or the division of scores for each part with correct and incorrect answers which is known as verification (McMillan, 2013). This kind of feedback focuses on the task or product and is used for accountability purposes. According to Boud (2000), feedback is not useful unless students can utilize it to enhance their work (Havnes et al., 2012). In summative feedback, even if students are provided with explanations, they are more concerned with their grades so they tend not to focus on improving their progress (Higgins, et al., 2001). In contrast, formative feedback is process-focused and given to students with the aim of changing their way of thinking or acting in order to enhance their learning (Shute, 2008). In formative assessment, feedback is not the end of the process. Students have various opportunities to revise their work (Heritage, 2010). The necessity of formative purpose of feedback is highlighted by Gigante, Dell and Sharkey (2011) who claimed that feedback can be effective when it reassures proficiency, directs future learning, encourages positive behaviour, indicates areas for growth and demands reflection.

To give effective feedback, teachers need to consider some principles. Effective feedback should be “goal-referenced, tangible and transparent, actionable, user-friendly, timely, ongoing and consistent” (Wiggins, 2016: 10-16). First of all, the feedback should be related to the learning goals and success criteria which have clear tangible standards of accomplishment, and align with course learning outcomes. A goal-referenced feedback explains how students meet assessment standards, identifies knowledge gaps, and addresses errors and misunderstandings

(Hatzia Apostolou, and Paraskakis, 2010). The transparency of feedback is similarly parallel to the qualities of rubrics and goals by making the feedback observable and meaningful (Wiggins, 2016). Thirdly, the quality of being actionable requires specificity, detail, and facilitation so that the students are motivated to act after getting the feedback (Gibbs and Simson, 2004). In addition to being actionable, effective feedback should be understood by the recipient easily and it shouldn't force the recipient to focus on too much information or have to improve too many areas at once (Brookheart, 2017). Timely feedback refers to giving feedback just before a student passes on to following tasks, at the time when the work is still fresh in their minds. It shouldn't be immediate or delayed for days or weeks (Wiggins, 2016). Ongoing feedback is a quality of formative assessment. While in summative assessment, the feedback is given at the end so that the students do not have a chance to work on the feedback and attain better results, in formative assessment feedback is given throughout the period of teaching to continuously make alterations to teaching and learning (Boston, 2019). Finally, consistency of feedback is about being consistent with the curriculum, learning goals and tasks. Students sometimes receive inconsistent feedback for the same performance from different teachers or from the same teacher. To provide accurate feedback that is measured by the same standards, feedback needs to be based on descriptive rubrics (Wiggins, 2016).

For many years, in educational contexts, feedback is perceived as a practice in which someone with greater knowledge gives advice to someone with less experience on how to improve upon what they did emphasizing the role of the teacher rather than the students. (Butler and Winne, 1995). With the *Assessment for learning*, students have been accepted as active participants in learning, not passive recipients and they have an ownership in and a responsibility for their own learning (Clarke, 2004).

In the traditional way of assessment, students serve as test subjects, supplying information to assess and the grading process is delegated to 'experts' with experience and they are hidden from students (Falchikov, 2004). When students do not have active participation in assessment, they do not benefit from it or they have a negative attitude towards assessment. Deakin-Crick and Harlen (2002) relate this to five factors. First of all, when assessment is about the decision of success and failure, if students do not do well, they get lower self-esteems and tend to give up or when

the focus is to fail or succeed, this might lead to memorizing the material in order to pass the test rather than truly learning the subject matter. Also, students feel anxious before exams due to lack of failure. In addition to that, they do not want to be judged by others according to their performances in a test. Finally, students believe that education that focuses largely on material that will be tested tends to limit their opportunities for learning.

Researches show that involving students in assessment fosters students' motivation, self-esteem and autonomy; increases higher achievement, and collaboration, lowers anxiety and creates a positive learning environment (Aslanoğlu, 2022; Birjandi and Tamjit, 2010). In classroom assessment, student involvement is greatly emphasized with AfL and AaL. Students are involved in assessment via self-assessment, peer-assessment, performance-assessments and portfolios.

In self-assessment, students participate in the process of creating the criteria and assess their own work by using it. They also direct their own learning by reflecting on the learning process. Incorporating student self-assessment might be a starting point for reducing the authority of teachers in the assessment process. This would empower students in the assessment process by allowing them to begin the feedback process by recognizing their strengths, weaknesses, and feedback requirements (Piper et al., 2019). When self-assessment is used in combination with teacher-generated feedback, it encourages students to take a more active role in their learning by giving a feeling of ownership and self-efficacy that transcends the limitations of conventional teacher-led assessment (Apridayani and Sakkanayok, 2024).

During self-assessment students use both internal and external feedback. External feedback can be gathered by asking teachers, classmates, etc. for their feedback or by monitoring your own work (e.g., including comparing one's own or others' past performance, examples, and criteria for assessment) (Yan and Brown, 2017). Only when students compare these with their own work and use them for internal feedback, they see changes in their performance or learning (Nicol, 2019). Self-feedback is unique in that it speaks to the individual learner, who acts as both a giver and a recipient of feedback (Panadero et al., 2019). On the other hand, students might provide subpar internal feedback that hinders learning. A lack of strategy knowledge may cause students to adopt ineffective strategies and techniques; lack of

motivation or time and resource constraints may cause students to reduce their goals without gaining additional knowledge; or students could overestimate their learning performance and cease studying due to poor self-assessment (Chou et al., 2015). Thus, self-assessment practices need to be planned well before being applied.

Peer feedback is a mechanism of communication that engages students in discussions about performance and expectations and it is mostly about thorough comments without official grading via using success criteria (Liu and Carless, 2006). Peer feedback is a useful method as it generates a lot of responses and exchanges amongst students and boosts confidence by fostering a friendly and cooperative discourse in which two-way feedback is developed (Rollinson, 2005). While peer-feedback is an inevitable part of formative feedback, some problems may occur during application. First of all, students may not have enough confidence in their ability to give meaningful and comprehensive criticism. In addition to that, when students perceive their partners as being low-competent, they may mistrust the given feedback. Lastly, when providing feedback, students may simply point out the weak parts of their peers' performance, which results in comments that are devoid of encouragement and full of criticism (Omeliicheva, 2005).

Compared to self-assessment, peer assessment is a more collaborative activity in which students learn from each other. During peer assessment, students judge other students' work critically by analysing the criteria and give constructive feedback by pointing out the areas of improvement (Hovardas et al., 2014). According to Zhao (2010), when students are engaged in discussion with their peers, commend each other's work and suggest improvements, they have a clear understanding of the feedback and thus they benefit from it more.

3. Methods of Classroom Assessment

As for the methods used in the classroom, it is possible to find many distinct general categories of assessment methods. In this part, assessment methods for reading, writing, listening, and speaking will be examined under three major categories: a) instructor-made assessment, b) student-conducted assessment, and c) standardized tests (Cheng et al., 2004).

a. Assessment of reading

Reading is a receptive skill that requires cognitive processes of

“comprehension” or “understanding” (Paris and Stahl, 2005). The most frequently used comprehension skills include skimming, scanning, finding the main ideas; identifying the text type; inferencing; distinguishing between fact and opinion; identifying the writer's purpose; drawing conclusions from the text's implications, and guessing unknown words (Hubley, 2018). Those skills are assessed both in classroom and standardized tests. Teachers generally use product-oriented questions such as cloze items, multiple choice, matching, true-false statements, fill-in-the-blank questions, and open-ended comprehension questions as they take less time, are highly practical, and are simpler to grade (Boubris and Haddam, 2020).

Standardized tests also have disadvantages while assessing reading. They are often given once or twice in a term and tend to assess lower comprehension skills in an isolated way and do not let students be a part of the assessment process (Valencia et al., 1994). Furthermore, they prevent teachers from supporting students by tracking students' progress, tailoring their teaching according to students' needs, and improving their instruction (Mkhwanazi, 2014). For that reason, they need to be used with alternative assessments in which students participate in the process such as reading-response journals, performance assessments, reading aloud, reading conferencing, and student portfolios.

One tool that students use to document their initial reactions to reading experiences is the reader response journal. Students first write journal entries following a section of a work that they have read. These can be a summary of the text, a reaction, or a reflection. Then they discuss it with their classmates or teacher. This way they can have a deeper understanding of the text (McIntosh, 2010). Performance assessments are also used as an alternative assessment in classroom assessment. After reading the text, students can work on tasks that are similar to daily life activities. In performance assessments, teachers can both guide students and assess their performance by observing and interpreting what they learn from the text and how they use this information in real life. Assessment can be done through checklists and rubrics (Afflerbach and Co, 2015). In the reading-aloud method, teachers/students read a text to students. During reading, teachers ask questions to students before, during, or after reading; encourage them to make guesses about what they are reading; model comprehension strategies, practice new vocabulary and allow students to reflect on the text (Ceyhan and Yıldız, 2021). Reading

conferencing is a part of one-to-one student conferencing where teachers meet students to discuss the texts selected by students. During the sessions, students are given opportunities to get feedback from the teacher, to consider the book they are reading in several ways through discussions, to set goals for future readings, to engage in meaningful conversations, and to develop /enhance strategies during reading (Russo, 2020). A portfolio assessment is a way of assessing students' progress via their work collected over a period of time. A reading portfolio may contain numerous items based on the learning objectives ranging from student-selected text to stories, news, articles, etc. but the core elements of the portfolios are assessment, responses observation, and reflection. For reading assessment, a piece of evidence such as running records can be used. Reading logs and story maps are used as written responses to the text. Teachers' observations can be conducted via classroom observation reports or anecdotes. Self-assessment or peer assessment checklists or reading reflection forms are useful tools for reflection (Hamp and Condon, 2000; Wolf and Runyan, 1996).

As can be seen above, for an effective classroom assessment, students' reading comprehension should be assessed by combining different methods as each method is limited in itself (Klingner, 2004).

b. Assessment of writing

Writing proficiency demands mastery of grammatical, oratorical, vocabulary, and intellectual topics, therefore assessing writing skills is a difficult procedure (Heaton, 1988).

There are two approaches to assessing writing. These are direct and indirect approaches. Indirect assessment assesses students' writing sub-skills and it is usually related to the accuracy at grammar, vocabulary, or syntax level. In indirect assessment, teachers commonly include multiple-choice items, matching items, true-false items, and editing a sentence or paragraph in a written work (Chen, 2016; Coombe et al., 2007). Direct assessment, on the other hand, corporates all aspects of writing such as content, accuracy, sentence order, and organization, and assesses students' communicative competence in writing. In direct assessment, students are required to have a written production like narratives or essays (Coombe et al., 2007). In classroom assessment, both indirect and direct are used by teachers.

In writing assessments, teachers also consider imitative, intensive (controlled), responsive and extensive performance. *Imitative performance* is basically used to assess students' ability to use mechanics of writing at early levels. Copying letters or words, writing dictated words in a close test, writing the words corresponding to the picture, filling in forms, and writing words for the given words or abbreviations are some examples of the tasks used for assessing imitative performance. *Controlled writing or guided writing* aims at identifying the use of proper words in the right context by using accurate grammar structures. Some assessment methods are dicto-comp, transformation of grammatical forms, and picture cue writing. In dicto-comp, a teacher reads a paragraph two or three times and then asks students to write it by recalling it from their memory. In grammar transformation, students change the grammatical forms (tenses, questions-statements, etc.) In picture cue writing, students either write short sentences or stories based on the pictures. *Responsive tasks* are used to assess writing at a restricted discourse level through summaries, replying e-mails or descriptions of a short event. *Extensive writing* is the least guided writing assessment task among all types which consists of longer forms of written works such as essays or thesis and students need to pay attention to all the aspects of writing assessment like coherence, cohesion, organization, accuracy, etc. (Brown, 2004). Essays help students perform better because they force them to grasp complicated concepts broadly and look for connections rather than just memorizing information (Jacob, 2010).

In the past few years, debates regarding multiple-choice base and writing samples have surrounded writing assessment. While the defenders of multiple-choice-based assessment regarded it as reliable and easy to assess, those who support writing sample tasks claim that identifying errors in a text is not enough since students need to show their knowledge and skills by creating their own text. However, either using writing samples that are produced spontaneously under time constraints in response to a given topic together or multiple-choice questions does not also offer a solution to represent the writing performance (Camp, 1993). The main reason is that in the EFL context, teachers who assign their students subjects, give little feedback on grammar and vocabulary, and react to their final products. The assessment of student writing is done at the end and teachers' comments mainly serve summative functions (Lee, 2007). Formative assessment closes this gap by

incorporating feedback through self-assessment, peer-assessment, and student-teacher interaction. (Graham, , 2011).

Another significant issue with writing assessment is scoring. According to the different purposes, analytical and holistic scoring are widely employed by teachers. An analytical rubric has a range of categories, such as range and accuracy, task achievement, organization, etc. and the rater assesses a student's performance by giving points to each section (Çetin, 2011). By giving a detailed picture of writing performance in multiple areas, it helps teachers (or students) to see the strong and weak aspects of the work and thus facilitate formative assessment and feedback (Jönsson et al., 2021; Weigle, 2002). On the contrary, holistic scoring does not assess different aspects of the writing performance and teachers give an overall score for the whole work. In this respect, it is not complicated and takes less time which makes it more useful for large-scale assessments like placements (Ghalib and Al-Hattami, 2015; Metruk, 2018). Apart from their advantages, both scoring methods have their pitfalls. While analytic scoring can create mental fatigue in the assessors, holistic scoring is less reliable as the scoring is more subjective, especially in the case of multiple assessors (Zhang et al., 2015).

c. Assessment of listening

Listening is an important receptive skill that is highly necessary for communication in daily life (Kassem, 2015). Homouda (2013) also emphasizes its significance in terms of understanding the spoken input which is vital for learning. At the same time, apart from the listening process itself, several elements such as the speaker, the message, and the listener make listening a difficult skill (Brown and Yule, 1983) As the listening process is complex and not possible to observe directly, assessing listening comprehension is needs to be considered carefully.

Rost (2013) identifies the six most commonly used methods employed for assessing listening which are discrete point, task-based, integrative, communicative, interview, self-assessment, and portfolio. In *discrete point assessment*, language competency is thought to be achieved if the person understands discrete items that are combined to form language. The primary method of measuring listening comprehension involves giving test takers an utterance in the target language and determining whether they comprehend it or not. The two primary ways to do this are by giving true-false and multiple-choice questions (Hemmati and Ghaderi, 2014).

Task-based assessment involves tasks that require responding appropriately with nonverbal cues after listening. The closed task with only one possible answer (e.g. selecting the correct date after listening or gap filling) or an open task that requires filling in the gaps with multiple words are sample tasks for task-based assessment (Rost, 2013). *Integrated tests* require test-takers to use at least two language skills. Integrated listening-to-summarize tasks combine listening and speaking or listening and writing skills as the test-taker needs to provide a written or oral summary or description following the listening (Rukthong, 2021). In addition to summarizing and description, dictation and note-taking are other integrated tasks that combine listening and writing. In dictation, students either complete notes using words or phrases or write what they hear at the same time. Dictating has a positive impact on improving writing and spelling as well as hearing (Syakur, 2020). Note-taking is noting the key information during listening. Apart from being an assessment method, it is a skill that people use to recall information in their daily lives (Hayati and Jalilifar, 2009). *Communicative methods* focus on the communicative aspect of language and hence have a pragmatic purpose. The examples may be written tasks (e.g. writing an advice letter after listening to a problem) or oral tasks e.g. (following directions) (Rost, 2013). In communicative tasks, test-takers must comprehend the written or spoken source in order to complete the task successfully (Rukthong and Brunfaut, 2019). *Interviews* are also a type of communicative method so usually, they are assessed together with speaking skills. The teachers can interview students or two students can interview each other. Scores are given according to a rubric/checklist, based on how well the test taker answers questions (Rost, 2013).

Self-assessment and *peer assessment* are very helpful tools for students to be actively engaged in their learning process. Following listening activities, students can use checklists to assess their own performance or they can keep journals to set goals for the future (Rost, 2013). Finally, as mentioned before, listening is a skill that cannot be observed directly so students may not be directly aware of their performances. By providing students with tangible evidence of their development,

Similar to what Rost (2013) suggests, assessment tasks can be classified into four categories as intensive, responsive, selective, and extensive. These are demonstrated in Table 2 (Ayuanita, 2013: 119).

Table 2. Assessment Methods in Listening

Intensive Listening tasks	Distinguishing phonemic pairs (Ex. grass-glass) Distinguishing morphological pairs (Ex. miss-missed) Distinguishing stress patterns (Ex. I can go-I can't go) Paraphrase recognition (Ex. I came from Taiwan-I am Taiwanese)
Responsive Listening Tasks	Repetition (students repeat a word) Multiple choice, open ended responses Simple discourse sequences Tough test
Selective Listening tasks	Listening cloze (students fill in the blanks) Verbal information transfer (students give multiple choice verbal response) Picture cued information transfer (students choose a picture) Chart completion (students fill in a grid) Sentence repetition (students repeat stimulus sentence)
Extensive Listening tasks	Dictation (students listen, usually 3 times, and write a paragraph) Dialogue (students hear dialogue - multiple choice comprehension questions) Dialogue (students hear dialogue - open ended response) Lecture (students take notes, summarise, list main points, etc.) Interpretive tasks (students hear a poem - interpret meaning) Stories - Narrative (students retell a story)

The following section will focus on the methods used for the assessment of speaking.

d. Assessment of speaking

The modern world requires students to have a good command of communication. Also, the communicative language approach puts a growing emphasis on speaking as a component of one's communicative competency. Thus, speaking is accepted as the most essential skill for learning a foreign or second language (Göktürk, 2016; Rao, 2019). According to Kürüm (2016) to have speaking competency, it is not enough to have grammatical accuracy, as speaking encompasses mechanics, functions, pragmatics, and social interaction. This makes speaking a difficult skill for students. As opposed to reading and writing, interactions in spoken communication take place simultaneously. A person starts the conversation by speaking about certain subjects. Then the respondent listens to his/her words, attempts to understand them, and then answers in the order that they occur (Ahmed, 2018). This aspect of speaking creates problems in terms of validity and reliability during assessment because the assessor needs to listen to the student while talking

and recall what s/he says to evaluate the performance and give feedback (Madsen, 1983; Pawlak and Waniek-Klimczak, 2015). Since speaking requires the use of multiple abilities, which develop at a different pace, it is not easy to define valid assessment tasks and criteria (Harris, 1969). Similarly, considering assessing speaking depends on the assessor's partial opinion, it can create reliability problems. To overcome this, in the case of summative assessment, the test should be scored by two separate assessors who are unaware of each other's scores (Karim and Haq, 2014). Furthermore, due to the elusive nature of speaking, speaking assessment in the classroom necessitates a holistic approach which is adopting multiple methods and various agents including the teachers and the students (Correria, 2016) as well as real-life contexts (Weir, 1990).

Generally speaking assessment tasks are divided into different categories by various authors. Luoma (2004) classifies them *open-ended tasks and structured or semi-structured tasks*. Brown (2004) divides the tasks according to skill types as *imitative, intensive, responsive and interactive tasks*. Nakatsuhara, Khabbazbashi and Inoue (2002) categorize tasks according to the conditions in which the oral speech is produced and they are *elicited imitation tasks, independent tasks, integrated tasks and interactive tasks*.

Some of the common assessing task types of speaking are “reading aloud, conversational exchanges, picture description, oral interviews, short talk, group discussion and role-playing” (Heaton, 1988, pp 88-105). Other methods used to assess speaking are oral presentations, giving oral directions, oral descriptions of an event, summarizing, and retelling a story in his/her own words (Brown, 2004; Louma, 2004).

Reading Aloud: Reading aloud is an indirect method of assessing students' pronunciation skills without being expected to produce utterances (Huang, 2010). During reading aloud, students read some texts at different lengths and the teacher or the examiner assesses the student's pronunciation or fluency. As the activity does not allow students to use the skill productively, it may be more useful at low levels (Gibson, 2008). Reading aloud can also be used as a self-assessment technique. As a part of classroom- assessment students may record their voice while reading a text and analyze it by using a criterion in terms of mechanical features such as stress, intonation, fluency, etc (Aoki, 2014).

Conversational exchanges: In conversational exchanges, students are given situations and are expected to respond in an appropriate way. According to the level, the responses can be fully controlled, semi-controlled, short or long (Heaton, 1988).

Picture description: Picture description tasks necessitate that students provide a verbal depiction of either a single image or a narrative within a specified timeframe. Students are also asked to describe the picture(s) or state their opinions or give information based on the picture(s) given (Karmila and Jufrizal 2019). Harmer (2017) highlights that, for a valid assessment, the pictures should be selected according to the age and level of students and they need to be clearly recognized.

Oral interviews: The most valid way of assessing speaking is the direct way in which students engage in face-to-face interactions with the assessors or their peers in pairs or groups (Brooks, 2009; Qian, 2009). Therefore, unlike the previous methods, oral interviews are more interactive and they may assess a wide spectrum of speaking skills, from novice to native. During the interviews, students interact with one or two assessors on various topics for 10-40 minutes, depending on their competence level. While reading aloud, conversational exchanges and picture descriptions have low-level validity, oral interviews demonstrate a substantial degree of reliability and validity. Though some consider one-to-one interviews far from being authentic as there is a hierarchy between the assessor and the test-taker which prevents them from forming natural conversations, they are accepted as valid because they assess speaking skills by requiring students to engage in real-world conversations. They are also reliable as two raters rate the same student's performance with a rubric (Ahmadi and Sadeghi, 2016). Furthermore, pair or group tasks can be included in one-to-one interviews. Interviews are also not limited to one task type. It is possible to use a lot of different task types at the same time. In an oral interview, there may be "description tasks, narrative tasks, instruction tasks, comparing and contrasting tasks, explaining and predicting tasks, and decision tasks" (Hatipoğlu, 2021: 158).

Small talks or oral presentations: In small talk tasks, students are asked to give prepared or unprepared brief talks about topics that are selected by themselves or the assessors. An example of it can be speaking tasks in which students are given one minute to get prepared and give a short talk for one or two minutes on the topic. To decrease student's anxiety and create a more welcoming environment, other

students may ask questions or interact with the student who does the talk. (Heaton, 1988). Oral presentations can be used in guided and free forms at any proficiency level (Farabi et al., 2017). Although oral presentations help students become more proficient in using languages, especially when it comes to speaking, developing critical thinking abilities and building confidence since they have to talk in front of the class (Ati and Parmawati, 2002), assessing oral presentations takes a lot of time and make students anxious (Mak, 2011). Also, students tend to memorize what they are going to say. To overcome these challenges, teachers are advised to teach some skills (e.g. presentation skills, organizing ideas etc.), set time limits, and prepare criteria and guidelines and share them with their students (King, 2002).

Small group discussions: For small group discussions, four or five students come together as a group to fulfil a predefined aim (discussing an even or coming to an agreement about a topic or solving a problem etc.) Group discussions foster meaningful and purposeful communication. They also help improve student's skills thanks to the monitoring and feedback of the teacher (Crisianita and Mandasari, 2022). Furthermore, Fulcher (1996) claims that when students have a group discussion with their peers, they express what they want and they have opportunities to actively engage in speaking when compared to being examined by a teacher.

Role-plays: Role-plays are activities in which students can have various social roles and practice speaking in different circumstances by placing themselves in someone else's position (Purnamawati and Suhartono, 2015). Role-plays are useful tools to teach and assess students' language use in different socio-cultural contexts (Liu and Ding, 2009). Conversely, role-play activities are time-consuming, making it challenging for teachers to manage all the students during class time. In a big class, assigning students to work in groups made it impossible for teachers to assess them individually when they were given different roles at the same time (Islam and Islam, 2013).

Retelling and Summarizing Tasks: This task is an integrative form of classroom assessment, combining both listening and speaking skills. After listening to their teacher, peers or a recording, students are asked to retell or summarize what they heard with their own words (Hirai and Koizumi, 2009). The benefits of story retelling and summarizing tasks are that they increase interaction between students and contribute to forming language complexity and increase accuracy (Safei, 2020;

Susilo and Husin, 2016).

Oral directions: Oral directions are parts of real-life instruction tasks whose basic aim is to convey the information and ensure that it is comprehended. This usually means brief interactions between the speaker and the audience (Luoma, 2004).

In addition to these methods, self and peer assessments are used to assess speaking performance in the classroom setting. While employing these methods in the class, there are some procedures that need to be followed. Before assessing the performances, teachers need to prepare criteria and go through each item to clarify the points. Rather than assigning grades, criteria should be used to identify the areas of strengths and improvements. Teachers can start assessing weak and strong samples before letting students assess themselves. Since it is difficult to remember, it may be helpful to ask students to record their own performances. When students finish assessing themselves/their peers, teachers need to come together with students and provide feedback on how to improve their performances (Başak, 2019)

4. Teachers' Beliefs about Classroom Assessment Practices

Teachers' beliefs are mental constructs that are based on personal, subjective truths. They are held by experiences and they affect teachers' judgments, behaviours during instruction (Opre, 2015). Teachers' beliefs on teaching and learning affect what they teach, how they teach it and how they interpret assessment (Munif et al., 2019); their classroom behaviour, teaching style, and approach to the students and their "goals and roles" (Li, 2012).

There are many internal factors that shape teachers' beliefs such as their learning experiences, their teaching experiences, their abilities and character, and their education and culture (Gilakjani and Sabouri, 2017). In addition to these, external factors also influence teachers' beliefs like students' motivation and proficiency levels, the assessment policy of the school and the class size (Cheng et al., 2004).

Several studies investigated the beliefs and practices of EFL teachers regarding to classroom assessment from different perspectives both in Turkey and abroad (Boubris and Haddam, 2020; Cheng et al., 2004; Korkmaz, 2021; Munoz and Escobar, 2012; Latif and Wasim, 2022; Önalın and Karagöl, 2018; Saad et al., 2013;

Widiastuti, et al., 2020).

Some researchers were interested in teachers' beliefs related to assessment procedures and difficulties. Mutlu (2022) in her MA thesis investigated Turkish EFL teachers' beliefs, their assessment procedures, and difficulties encountered during the assessment process. The findings of the research showed that teachers had different beliefs regarding using traditional and alternative assessment methods to assess four skills and their beliefs were affected by students' level of language, teachers' lack of assessment knowledge, educational policy, the absence of technological tools and type of school.

Other researchers focused on the changing beliefs of teachers according to experience, nationality and education context. A study conducted by Bural (2021) for instance, explored the Turkish and Spanish EFL teachers' beliefs on the possibility of applying dynamic assessment in relation to the level of degree and gender and found that both Turkish and Spanish teachers had positive beliefs toward the suitability of dynamic assessment and there was no difference for the beliefs according to the gender and educational level.

Similarly, Çakır (2020) conducted a comparative research. She researched foreign language teachers' beliefs and practices in primary schools and tried to find out whether the quality of assessment matched the real practices of teachers in Turkey, Finland and Italy and she found that the beliefs of teachers in Finland, Italy, and Turkey indicated an inclination to employ more formative, student-centred assessment methods. Generally, teachers' beliefs regarding content-skills, exercises, and assessment methods appear to be closely aligned with their practices to some extent.

Özdemir-Yılmaz and Özkan's conducted a study on the tertiary level (2017) and explored the methods and different beliefs of teachers working at public and private universities. The data revealed while the teachers mainly used student-centred methods, for diagnostic purposes for assessing listening, speaking and writing, the use of instructor-made methods was also noticeable in all skills. However, their feedback process tended to be lacking of being constructive and process-based. Besides, they were inclined to use standardized exams. The researchers concluded that the dependence of the exams could be related to the washback effect of proficiency exams and their practices were not in line with the formative assessment

because of their previous habits of using mostly summative assessment. Also, the data did not reveal any significant difference between the beliefs and practices of teachers working in public and private universities. Additionally, Önalın and Karagöl (2018), searched the relationship of EFL teachers' beliefs on assessment with the undergraduate departments and experience in teaching. According to the results, teachers had strong beliefs on using formative assessment and self-assessment and teachers' experience or undergraduate department had no effect on their choice of assessment types.

There are also studies about teachers' beliefs on classroom assessment abroad. These studies focused on the teachers' beliefs together with the purposes and methods. Van, Tran and Tran (2021), studied teachers' beliefs and classroom assessment practices in Vietnam concerning the purpose of classroom assessment. After analysing in-depth interviews of six teachers and testing documents, the results showed that with the effect of high-stake exams, teachers believed that the main purpose of classroom assessment was to prepare them for the exam so they paid more attention to linguistic elements in the content and summative test formats. Widiastuti et al. (2020) also wanted to learn the correlation between formative assessment procedures and instructors' beliefs in EFL classes in terms of CPD (Continuing Professional Development) participation. The study's findings demonstrated that instructors who engage in more CPD activities had stronger beliefs about formative assessment than teachers who engage in less CPD activities. On the other hand, it was not reflected in their classroom practices because of the intensity of their CPD involvement and limited time.

A recent study by Rahmati and Nushi (2023) investigated the importance of students' views, the distinction between high-stakes and low-stakes tests, and the function of computer literacy in connection to justice and fairness in technology-enhanced classroom assessment during Covid were discussed beliefs. The study indicated that teachers considered that the views of learners were a useful source of information on fair and just classroom assessments and that justice and fairness were both significant in assessments with high and low stakes. Furthermore, the teachers stated that thanks to technology, they learned using paper assessments together with online assessments.

Wang, Lee and Park (2020) undertook a study with 136 EFL teachers in

China to examine the degree to which university instructors' ideas regarding classroom writing assessment are consistent with their self-perceived practices, as well as the reasons that lead to mismatches between beliefs and practices. Similar to other studies' results, this study showed that even though teachers believed Assessment *for* Learning helped students to be more responsible in their learning process with writing than Assessment *of* Learning, they relied more on Assessment *of* Learning during classroom teaching. Moreover, it was discovered that teachers were prevented from acting in accordance with assessment beliefs by a variety of factors, such as inadequate training in language assessment, student characteristics, work constraints, and the instrumental value of high-stakes tests. Finally, more experienced teachers tended to be more inconsistent with their practices and beliefs.

Lastly, Muñoz, Palacio, and Escobar (2012) studied EFL teachers' beliefs on assessment in a university in Colombia and the findings pointed to a gap between the participants' stated beliefs and practices, indicating the need for teachers to have additional opportunities for self-reflection self-evaluation, and formative assessment techniques assistance.

Despite extensive research on teachers' assessment beliefs and practices, there remains a consistent gap between what teachers believe to be ideal and what they actually perform in their classrooms. Therefore, this study aims to explore the true factors contributing to the gap by investigating the purposes, methods, beliefs and practices of EFL instructors in Turkey about classroom assessment.



III. METHODOLOGY

The scope of this chapter is the methods and procedures utilized in the study which consists of the research design, participants and setting selection, data collecting methods, tools, and data analysis process. To investigate teachers' purposes, methods, beliefs and practices of EFL Instructors about Classroom Assessment at Turkish Universities, three research questions are formulated:

- 1- For which purposes do EFL Instructors at Turkish universities use classroom assessment?
- 2- Which methods do EFL Instructors at Turkish universities use to assess four skills (reading, writing, listening, and speaking)?
- 3- What are EFL Instructors' beliefs about Classroom Assessment? To what extent are the beliefs of instructors in line with their practices.

A. Research Design

The aim of this study was to explore purposes, methods, beliefs and practices of classroom assessment used by EFL Instructors working at the preparatory schools of Turkish universities.

To collect reliable data, a mixed-method study approach was adopted. The mixed method is an empirical approach that integrates both qualitative and quantitative data. Mixed-method research leverages the strengths of both qualitative and quantitative methods by minimizing their drawbacks and providing a comprehensive overview of the issue (Punch, 2005). Moreover, it allows us to draw stronger conclusions since the data are examined from several angles. Both approaches can offer more breadth and depth, and when combined, they validate or enhance one another (Pole, 2007). As the examination of classroom assessment needs multiple perspectives (numerical data related to the frequency of the use of the method, explanation about instructors' beliefs, methods and purposes of classroom the assessment, classroom observations to provide evidence for their practices), the

mixed method is the most suitable design for this study.

For collecting quantitative data, the survey method was employed. Surveys are helpful tools when the researchers need to gather data about the demographic information or views of a large population (Ponto, 2015). When the researcher's objective is to gain a deeper grasp of the participant's distinct perspective rather than a generalized comprehension of phenomena, semi-structured interviews are the preferred way of gathering data (Adeoye et al., 2021). Since this research aims to learn about teachers' views and practices, for the qualitative data, semi-structured interviews can help researchers to compare teachers' interview results with their actual practices, to better understand the challenges faced by teachers and to look into the possibility of future improvements (Chen, 2006).

B. Setting and Participants

The setting of the present study is preparatory programs at state and foundation universities in Turkey. In both types of universities, the standard summative assessment as well as the procedures of formative assessment is mostly governed by Testing and Assessment Units. At the beginning of the term, students are grouped according to their levels and after completing a level, they move on to the other level. At the end of the year, if they pass the proficiency exam which is prepared and conducted by testing units, students start their department. While teachers invigilate students and grade exams, they do not see the standardized exam questions before or the responsibility of standardized tests methods lies only with testing unit members and other instructors have no saying or involvement with any assessment process for the standardized test. However, in classroom assessment the instructors have flexibility and their methods are not limited to those that are used in standardized exams.

In the present study, 100 EFL Teachers working at the preparatory programs of 31 foundation and state universities in Turkey were selected through convenient sampling. Of 100 participants, 79 were females and 21 were males. For the purpose of gathering demographic information about the participants, they were asked to provide information about their gender, age, years of teaching experience and educational qualifications, assessment training, and perceived knowledge about language assessment and evaluation.

Table 3. Demographic Characteristics of Questionnaire Participants

Variables	Number of Total Participants	P	Percentage %
Gender			
Male	21		21
Female	79		79
Age			
18-25	6		6
26-30	23		23
31-35	39		39
36-40	18		18
41-45	7		7
46-50	3		3
over 50	4		4
Years of Teaching Experience			
less than 5 year	28		28
6-10 years	42		42
11-15 years	17		17
16-20 years	7		7
more than 20 years	6		6
Educational Qualifications			
BA in English Language /Field Teacher	27		27
BA in English Language and Literature	23		23
MA	44		44
PHD	6		6
Assessment Training			
None	39		39
A course	36		36
A workshop	25		25
Perceived Knowledge about Language Assessment and Evaluation			
No Knowledge	1		1
Poor Knowledge	4		4
Good	67		67
Very Good	20		20
Excellent	8		8

The number of female respondents (n=79) were higher than male respondents (n=21 22). In relation to age, 39% of participants were between 30-35 years old; 23% of the participants were between 26-30 years old; 18 % of the participants were between 36-40 years old; 7% of the participants were between 41-45 years old; 6% of participants were between 18-25 years old; 4% of the participants were over 50 years old; 3% of the participants were between 46-50.

The majority of the participants 72% had more than 5 years of experience while 28% of participants had between less than 5 year of experience. The

participants who had 6-10 years of experience were 42%. The participants with 11-15 years of experience were 7% of participants had 16-20 years of experience and 6% of participants had more than 20 years of experience.

Regarding educational qualifications, most participants (n=44) held Master's degrees. Following those, 27 participants had BA degree in English Language/Field Teacher; 23 participants had BA degree in English Language and Literature and 6 participants had PHD degree.

Regarding assessment training, 39% of the participants didn't have any training while 36% of participants took a course. Those who attended workshops relating to assessment were 25%.

Concerning perceived proficiency in language assessment and evaluation, 67 participants affirmed possessing a good level of knowledge; 20 participants said that they had very good knowledge; 8 participants mentioned that they had excellent knowledge; 4 participants asserted that they had poor knowledge and 1 participant expressed that s/he had no knowledge about language assessment and evaluation.

To sum up, the majority of the participants (n=79) were female, around the age of 30-35 (39 %), with a MA degree (44 %) and with 6-10 years of experience (37.9%). Most of the participants have no assessment training (39%) but overall more than half of the number of participants (67%) perceive their knowledge of assessment and evaluation as good.

C. Data Collection Instruments

This chapter gives an in-depth description of data collection tools, which are a questionnaire, interview questions and observation checklist.

1. Questionnaire about Teachers' Views, Purposes and Methods of Classroom Assessment

In this study, to investigate the beliefs and practices of EFL Instructors working in preparatory classes at Turkish universities, a mixed method was adopted. First, survey research was used. The study was conducted by using a close-ended questionnaire (Appendix A) which was designed by Swai (2019).

The scale is divided into four sections which consists of 57 items in total. In the second and third part, the participants stated their views using a 5-point Likert scale: a scale with five options: 1 (Never 0%), 2 (Rarely 25%), 3 (Often 50%), 4 (Usually 75%) and 5 (Always 100%). In the fourth part, they responded using a 5-point Likert scale: (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree.

In the first part, there were six questions asking the demographic information of the participants (gender, age, years of teaching experience, education qualifications, courses or workshops the participants attended on assessment and evaluation, and participants' perceived knowledge about language assessment and evaluation). The second part consisting 11 items focus on purposes of the assessment, which were divided into pedagogical (student), pedagogical (instructional) and administrative purposes parts. The third part of the questionnaire includes 37 items providing the information about the type of methods to assess reading, writing, listening and speaking. Each part (reading, writing, listening/speaking) consist of three sub-headings: (1) instructor made, (2) student-conducted and (3) standardized test. In the last part, there are 8 items collecting information about the participants' beliefs about language assessment. While the first item is related to beliefs toward summative assessment, the other 7 items are about beliefs toward formative assessment. The last question of this part is included to identify the methods the respondents use to provide feedback.

Although the original questionnaire's reliability was ascertained by the author (purposes= .867, methods= .931, and beliefs= .713), considering that the "reliability is greatly affected by the sample characteristics" (Bademci, 2006: 7), and as the characteristics of EFL teachers in Turkey may be different from the Jordanian EFL teachers, the internal consistency of the items was tested again for each category. The questionnaire was proved to be reliable with the values of purposes of assessment= .831 (good), methods of assessment= .952 (excellent) and beliefs=.823 (good).

2. Semi-Structured Interviews

Semi-structured interviews were conducted in English with 10 participants who had already filled in the survey so that they were able to share their experiences and views more freely related to the practices and methods. 10 interview questions

related to the practices and beliefs of EFL Instructors were formed according to the findings of questionnaire results to clarify some answers / results. The interview questions were based on the items in the questionnaire by Swai (2019).

D. Data Collection Procedure

Before starting the interviews, the researcher sent an e-mail to Malak Swai who designed the questionnaire and interview questions, and got permission to use the questionnaire. Upon getting the consent, the instruments then were sent to the Institute of Educational Science's Ethics Committee at Istanbul Aydın University in order to get the required research approval for gathering data with these tools. When all the consents were obtained, the questionnaire was transferred to Google Forms. While preparing the questions, all the questions were made compulsory to answer in order to prevent the respondents to skip any items. At the questionnaire, the participants were informed about the scope and aim of the research ensured that their responses would become anonymous and they had right to withdraw from the research any time and they would be informed about the results if they requested. Then, the link was to participants at different universities in Turkey via e-mails and WhatsApp groups and volunteer participants completed the questionnaire. The data collection process completed in one month in 2024 and none of the participants withdrew from the research.

10 interviewers which makes 10% of instructors who filled in the questionnaire, were selected by random sampling. Instructors were requested to attend the interview via e-mails and all of them accepted the request. Since instructors were living in different cities, the interviews were conducted via MS Teams platform online. Before starting, the researcher informed the interviewers about the purpose of the research and were requested to provide as detailed answers as possible, emphasizing there were no right or wrong answers regarding their classroom practices. Also, they were ensured that the data would not be shared by anyone and their identities would kept anonymous. The interviews which, took approximately half and hour, were conducted in English as the participants were all EFL instructors. Following the interviews, the data was transcribed and analysed thematically.

E. Data Analysis

The data collected from the questionnaire was transferred into SPSS version 26. The data was exported into SPSS directly from Google forms excel sheet to reduce data entry errors and was analysed with descriptive statistics (mean, median and standard deviation and frequency). Each section was analysed separately.

For the analysis of the qualitative data, content analysis was applied to the semi-structured interviews. For this purpose, the interviews employed in English were recorded and transcribed. After that, the data was coded and categorized according to the contents. Finally, the categories were grouped according to each question as themes.

The following chapter was devoted to the findings and discussion of the results of the current study.

IV. FINDINGS

This study aimed to investigate purposes, methods, and beliefs of EFL Instructors who work at the preparatory schools of Turkish universities, regarding classroom assessment. To investigate teachers' purposes, methods, beliefs and practices of EFL Instructors about Classroom Assessment at Turkish Universities, three research questions are formulated:

1- For which purposes do EFL Instructors at Turkish universities use classroom assessment?

2- Which methods do EFL Instructors at Turkish universities use to assess four skills (reading, writing, listening, and speaking)?

3- What are EFL Instructors' beliefs about Classroom Assessment? To what extent are the beliefs of instructors in line with their practices.

The section is divided into three parts: (1) the findings regarding the purpose of EFL instructors in terms of classroom assessment, and (2) methods of classroom assessment and (3) the beliefs of Turkish Instructors on classroom assessment. Each section presents the questionnaire and interview results for the related area.

A. Findings Regarding the Purposes of Classroom Assessment

The first research question aimed at investigating the purposes of EFL instructors in terms of classroom assessment. This part focuses on the findings from the questionnaire and interviews.

1. Findings from the Questionnaire

In the first part of the questionnaire, the participants responded to a set of items about the purposes of classroom assessment which was divided into pedagogical (student-centred), pedagogical (instructional purposes) and administrative purposes parts and had 11 items in total. The data are sorted in descending order, from the highest mean to the lowest.

Table 4. Purposes of Assessment as Stated by Instructors

Purposes	Mean	Median	Mode	SD*
Student-Centred Purposes				
To provide feedback to my students as they progress through the course	4,47	5,00	5	,658
To diagnose strengths and weaknesses in my students	4,40	5,00	5	,725
To obtain information on my students' progress	4,20	4,00	5	,910
To determine the final grades for my students	3,98	4,00	5	1,172
To formally document growth in learning	3,85	4,00	4	1,158
To motivate my students to learn	3,83	4,00	5	1,120
To make my students work harder	3,77	4,00	4	1,043
Instructional Purposes				
To plan my instruction	3,82	4,00	4	,957
To diagnose strengths and weaknesses in teaching	3,80	4,00	4	1,082
To group my students for instructional purposes in my class	3,13	3,00	3	1,134
Administrative				
To provide information to the school administration	3,76	4,00	5	1,327

*SD = Standard Deviation

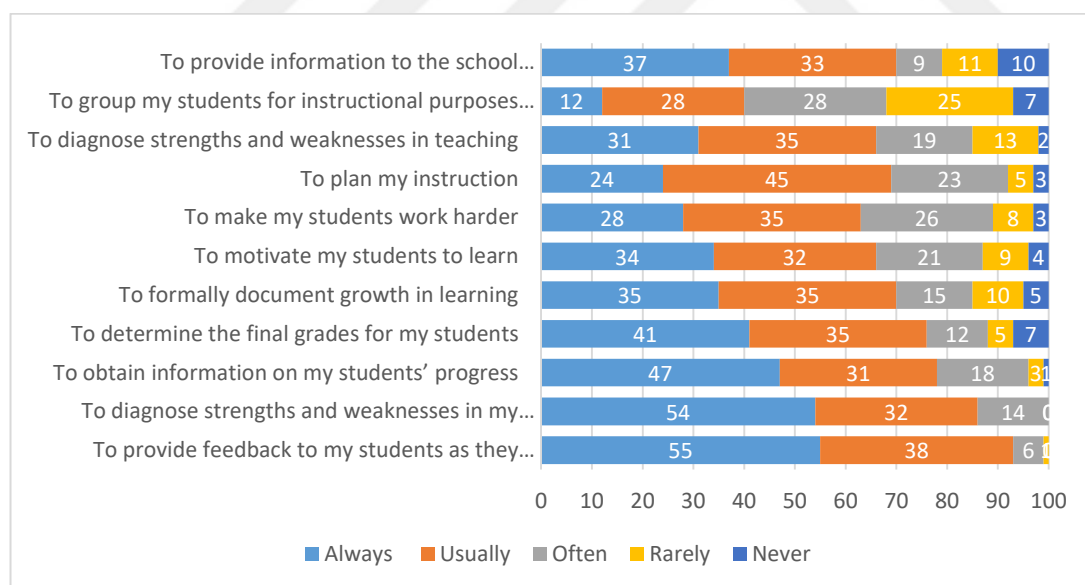


Figure 1. Percentages of the purposes of assessment

As Figure 1 illustrates, instructors use classroom assessment to fulfil a variety of different purposes. Most instructors (n=93) indicated that they use classroom assessment for student-centred purposes. They stated that they “Always” (55 %) or “Usually” (38%) use assessment to *provide feedback to their students as they progress through the course*. Following feedback, 86 instructors stated that they

“Always” (54%) or “Usually” (32%) use classroom assessment *to diagnose strengths and weaknesses in their students*. The third most frequently used purpose of classroom assessment by instructors was *to obtain information on my students’ progress* with 78 respondents in total selecting “Always” (47%) and “Usually” (31%). Other frequently used classroom assessment purposes which were “Always” or “Usually” used by instructors were: *to determine the final grades of the students* (76 %) and *to formally document growth in learning* (70 %). The results revealed the least indicated purpose was the teacher-centred purpose of *grouping students* with a percentage of 12% “Always”, 28% “Often” and 28% of “Usually” responses with the lowest mean score of $\bar{x}=3,13$ among all the items.

2. Findings from the Interviews

In order to learn about the purpose of classroom assessment, instructors were asked for which purposes they use assessment in the classroom. Upon analyzing their responses, a few recurring themes emerged. These themes were: to obtain information about their progress and learning, to provide feedback, to motivate them, to diagnose their strengths and weaknesses, and to group them. The most commonly used theme for the purpose stated was obtaining information about students’ progress. The following remarks of instructors make this point clear:

The purpose of classroom assessment is to check whether the students have the opportunity to obtain the classroom objectives or the syllabus-related objectives [to see] if they have met or not (Instructor 1)

Instructor 1 highlights the importance of classroom objectives as well as syllabus-related objectives. In formative assessment, assessment is regarded as a crucial component of learning. Thus, in order to obtain information about students’ progress, learning objectives play an important role.

Instructor 4 also shares the same purpose with the other two instructors by emphasizing that s/he uses classroom assessment to check if students understand the subjects or not. She also mentions they can use these results to guide students:

[Classroom assessment] helps teachers to see how well students understand the topic that we are teaching. Also, it helps students to know what they are doing and what they need to work on or the areas of improvement... (Instructor 4)

The second most common purpose was to provide feedback to students:

The most important one I'm talking about [is] to provide feedback to the students as they progress through the course. (Instructor 5)

According to Instructor 5, the purpose of using classroom assessment is to assist students during the learning process by giving feedback throughout the course rather than providing feedback at the end of the term. Instructor 3 also shares this belief. In addition to that s/he also mentions the significance of giving feedback for diagnosing their strengths and weaknesses and providing feedback for further needs:

I use classroom assessment to diagnose the students' learning and if they need assistance, I use it to provide feedback for further needs. (Instructor 3)

Instructor 9 provides us with information about the type of feedback and reveals the necessity of giving timely, instant feedback:

The main reason is to provide my students with feedback, and the feedback is like an instant one. I give them instant feedback and it makes them like learn what they need to do so they progress more smoothly. (Instructor 9)

Another common theme was diagnosing strengths and weaknesses. Four instructors stated that they use classroom assessment to diagnose strengths and weaknesses. One instructor emphasized the important of formative function of assessment:

We do not grade them for in class performance, so just actually [we us it] basically for [diagnosing] the strengths and the weaknesses of the students. (Instructor 2)

Diagnostic purpose of classroom assessment is further elicited by two other instructors who stated that they use classroom assessment to motivate students to learn rather than see what they learned. Instructor 1 acknowledges that the ultimate aim of assessment is not to test students' knowledge but to encourage them to learn more:

To contribute or to enhance learning process. The ultimate aim is

not to test students' knowledge, it's to encourage them. (Instructor 10)

Instructor 1 also acknowledges the significance of using assessment to motivate students and states the impact of assessment on students' active participation in the learning process:

And the second thing is motivating students and increase the involvement of each student [in learning process] (Instructor 1)

While it is clear from instructors' remarks that they use classroom assessment for student-centred formative purposes, they also stated that they use it for summative purposes for different reasons:

We use classroom assessment to group students according to their levels. Thanks to classroom assessment, they can just learn their levels (Instructor 6)

Classroom assessment, depending on the type of the assessment, serves formative and summative purposes. Our main purpose is to check our students' learning process like where they have been so far, and what can they achieve at the end of the course, like to see the outcomes, [if] they really achieve those outcomes or not. (Instructor 7)

The grouping purpose as expressed by Instructor 6, implies the necessity of using summative assessment to group students so that students will repeat the same level if they are not eligible or continue with a higher level. Instructor 3 and Instructor 8 describes the methods they use to group students:

It's more like, you know, placing students according to their levels. During the track, we don't send the students to other levels if it's a pre-intermediate level, they obtain some grades from certain quizzes, midterms and finals (Instructor 3)

For me it is evaluating students [performance], because after teaching and conducting the lesson, we need to see how they perform. The best way is for assessing either can be quizzes or it can be can do some midterm or final exams. (Instructor 8)

While the other two instructors seem to have no problems with the summative function of classroom assessment, one of the instructors disapproves of the value of

grading practices.

Actually, I have to [grade my students] because of the, I mean institute, but it's not my preference (Instructor 5)

B. Findings Regarding the Methods of Assessment

The second research question was to explore the methods which EFL Instructors employ from assessing reading, writing, listening and speaking. This part focuses on the findings from the questionnaire and survey.

1. Findings from The Questionnaire

The second research question guided the respondent to determine the methods of assessment to assess four skills (reading, writing, listening/speaking). In the study, the methods of assessment for the reading, writing, listening/speaking were divided into three categories: (1) instructor-made tests (2) student-conducted assessment and (3) standardized tests. Instructor-made assessment methods are those that are created and delivered by instructors and student-conducted assessments are the ones that include students directly in the assessment process (Cheng, 2004). The data are sorted in descending order, from the highest mean to the lowest.

Table 5. Methods of Reading Assessment

Methods	Mean	Median	Mode	SD*
Instructor-made				
Multiple-choice	3,91	4,00	4	1,006
True/false	3,82	4,00	4	,989
Matching items	3,82	4,00	4	1,029
Short answer items	3,67	4,00	4	1,064
Sentence completion	3,54	4,00	4	1,068
Cloze items	3,43	3,50	4	1,112
Student-Conducted				
Student summaries of what is read	3,41	3,00	3	1,055
Student portfolio	3,23	3,00	3	1,081
Peer assessment	3,32	3,00	3	1,043
Self-assessment	2,91	3,00	2	1,342
Read aloud	2,83	2,00	2	1,407
Standardized Test	3,68	4,00	4	,963

*SD = Standard Deviation

The data collected from the questionnaire pointed out that instructors utilize a range of assessment methods to assess students' reading proficiency. Table 5 lists the summary of the methods and the frequency of their use.

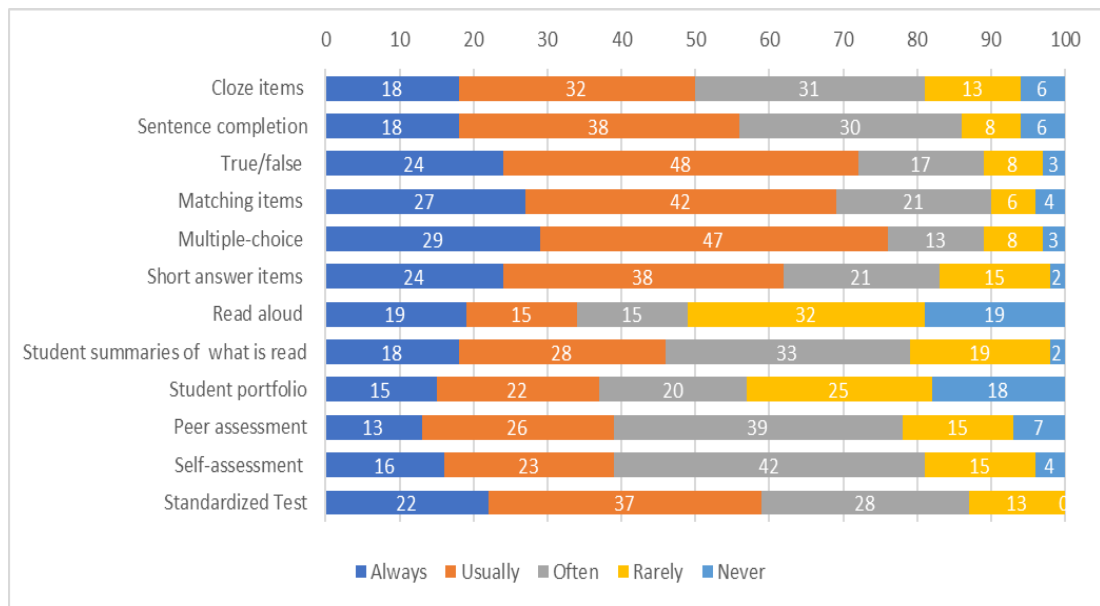


Figure 2. Percentages of the reading assessment methods

Instructor-made methods. According to Table 5, the most frequently used instructor-made methods were: *Multiple Choice items* (27% responded “Always” and 47% responded “Usually” and 13% responded “Often” while 11% responded “Rarely” or “Never”). The mean score (3,91) with a relatively low standard deviation indicates that the participants responded positively to both items. Other most frequently used method were: *True/False items* (24% responded “Always”, 48% responded “Usually”, 17% responded “Often” while 11% responded “Rarely” or “Never”) and *Matching items* (27% responded “Always”, 42% responded “Usually” and 21% responded “Often” and only 10% responded “Rarely” or “Never”) and. The less frequently used instructor-made method was *Close items* with 19% “Rarely” and “Never” responses.

Student-conducted methods. As it is referred from Table 5, instructor’s responses to student-conducted methods were less frequent when compared to their responses to instructor-made assessment methods (most participants responded “Often” or “Usually” rather than “Always”). While 61 % of respondents reported they “Usually” or “Often” use *Student summaries*, nearly a quarter of respondents (21 %) reported they “Rarely” or “Never” use them. The least frequently responded item was *Read read-aloud* method. More than half of 100 respondents (n=51) said they “Rarely (32%)” or “Never” (19%) use it. Similarly, while 15 % of respondents reported that they “Always” use *Student Portfolio*, 43 % of them responded “Rarely” or “Never” use it. Likewise, many respondents indicated they “Rarely” or “Never”

use *Self-assessment* (22%) or *Peer-assessment* (19%).

Standardized testing. More than half of instructors (59%) responded “Always” or “Usually” or “Often” for *Standardized tests* and no respondent selected “Never” response.

The degree of variety in the preferences or practices of teachers is reflected in the standard deviation figures. A higher standard deviation indicates more variation in the use of the method. The small standard deviation number (0.963) suggests that instructors are employing standardized assessments in a more homogeneous manner. The significant standard deviation figure (1.407) indicates that instructors’ usage of the read-aloud method varies significantly, perhaps due to different attitudes or preferences. On the other hand, the standard deviations of the student portfolio and self-assessment methods are marginally greater (1.081 and 1.342, respectively). This shows that instructors’ use of these two methods is more diverse.

Table 6. Methods of Writing Assessment

Methods	Mean	Median	Mode	SD*
Instructor-Made				
Short essay	4,12	4,00	5	1,008
Editing a sentence/ a Paragraph	3,85	4,00	5	1,201
Long essay	3,66	4,00	4	1,281
Multiple-choice items to identify grammatical error(s) in a sentence	3,05	3,00	3	1,306
Matching items	2,91	3,00	2	1,349
True/false	2,76	2,00	2	1,415
Student-Conducted				
Self-Assessment	3,39	3,50	4	1,118
Peer Assessment	3,24	3,00	4	1,147
Student Portfolio	3,19	3,00	5	1,461
Student journal	2,54	2,00	1	1,275
Standardized Test	3,69	4,00	4	1,212

*SD = Standard Deviation

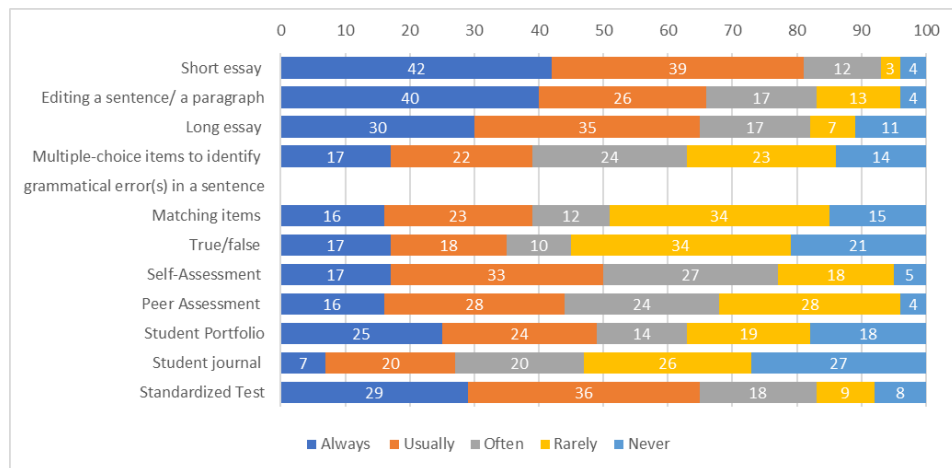


Figure 3. Percentages of the writing assessment methods

Instructor-made methods. As can be seen in Table 6, generally instructors employ each method with different frequencies. More than half of instructors (81) reported that they “Always” (42%) or “Usually” (39%) use *Short essay* to assess students’ writing. The following most frequently applied method was *Editing a sentence/paragraph* (66% responded “Always” or “Usually”). It was also observed that nearly half of instructors (n=65) use *Long essay* (82% responded “Always”, “Usually”, and “Often”). *True/ False* and *Matching items* were the least frequently used methods. For *True-False* method, 55% of instructors responded “Rarely” or “Never”. A similar finding was found out for *Matching items* and 49% percentage of instructors indicated that they “Rarely” or “Never” use it.

Student-conducted methods. The majority of instructors responded to student-conducted assessment methods less frequently with high percentage of “Rarely” and “Never” responds. Compared to the instructor-made assessment, the findings indicated that instructors of the sample used student-conducted methods less frequently than the instructor-made assessment. *Self-assessment* was the most frequent method and 50 instructors stated that they “Often” or “Usually” use it. On the other hand, nearly a quarter of the total number (n=23) of instructors stated that they “Rarely” (18%) or “Never” (5%) use it. The least frequent assessment method was *Student journal*; more than half of instructors (n=53) indicated that they “Rarely” (26%) or “Never” (27%) use it.

Standardized testing. Most of instructors (n=65) reported that they use *Standardized Tests* to assess writing skills. (18% selected “Usually). Only 8 instructors stated that they “Never” use it.

Table 7. Methods of Listening and Speaking Assessment

Methods	Mean	Median	Mode	SD*
Instructor-Made				
Multiple-choice test items following a listening Passage	3,99	4,00	5	1,133
Take notes	3,85	4,00	4	1,067
Prepare summaries of what is heard	3,59	4,00	4	1,111
Student-Conducted				
Oral interviews=	4,02	4,00	5	1,015
Dialogues				
Oral discussion with each student	4,00	4,00	5	1,064
Provide an oral description of an event or object	3,78	4,00	5	1,079
Oral presentations	3,73	4,00	4	1,109
Follow directions given orally	3,64	4,00	5	1,235
Give oral directions	3,62	4,00	4	1,204
Oral reading/dictation	3,40	4,00	5	1,435
Retell a story after listening to a passage	3,28	3,00	4	1,264
Self-assessment	3,23	3,00	3	1,262
Peer assessment	3,06	3,00	4	1,301
Standardized Listening Test	3,85	4,00	5	1,282
Standardized Speaking Test	3,76	4,00	5	1,349

*SD = Standard Deviation

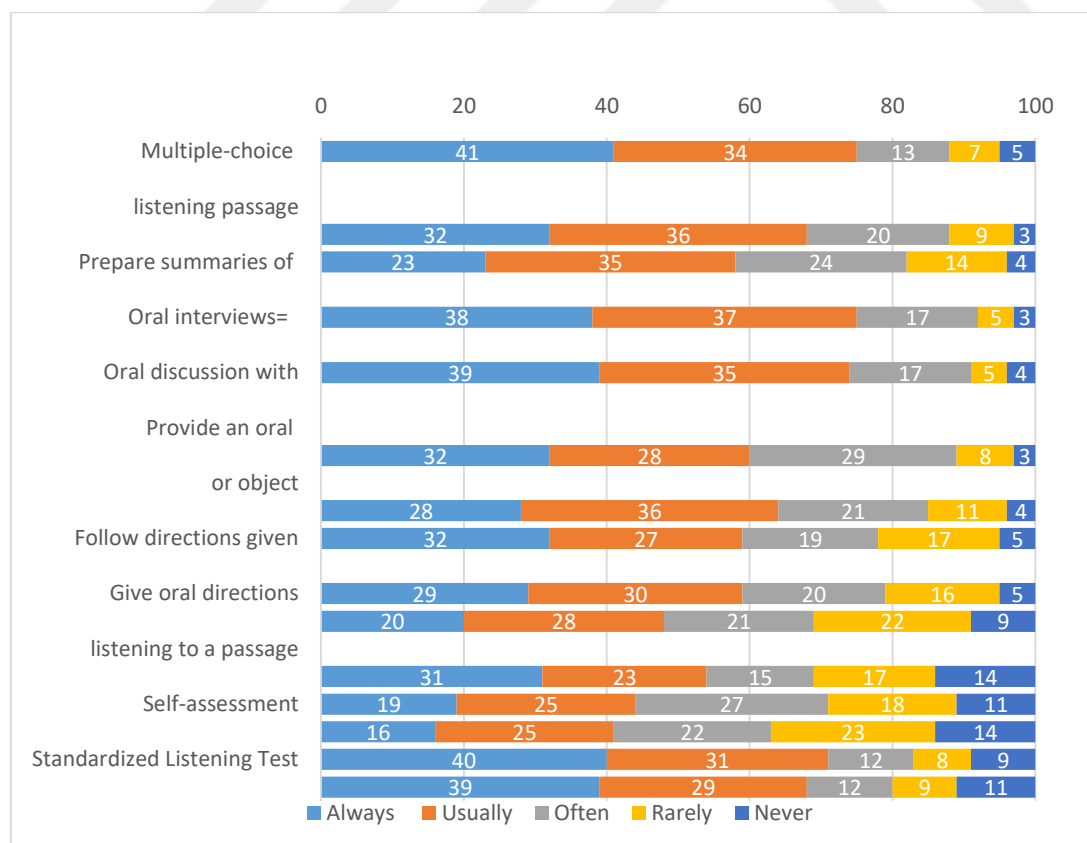


Figure 4. Percentages of the listening and speaking assessment methods

In this study, the data for listening and speaking methods were not collected and analyzed separately. As illustrated in Table 7, the full list of assessment methods for both listening and speaking as well as the percentages of instructors that selected each method.

Instructor-made methods. As can be seen from Table 7, *Multiple-choice Test Items Following a Listening Passage* was the most frequently employed method with $\bar{x} = 3,99$ and 41 out of 100 respondents stated that they “Always” use it and 34 respondents indicated they “Usually” employ it. Following that, the second frequently employed assessment method was *Take notes* with $\bar{x} = 3,85$. More than half ($n=68$) of the respondents indicated that they “Always” (32%) or Usually (36%) employ it. The least frequently used instructor-made method was *Prepare summaries of what is heard* with a percentage of 23% “Always” and 35% “Usually”. Overall, the participants responded to all statements in this category positively.

Student-conducted methods. Among all methods, the percentage of *oral interviews = dialogue* method was higher than the other methods with $\bar{x} = 4,02$. More than 50% of the respondents ($n=75$) responded “Always” (38%) or “Usually” (37%) to this method. This was followed by *Oral discussions with each student* ($n= 74$), and *Provide an oral description of an event or object* ($n=60$) When, the total percentage of “Rarely” and “Never” responses were considered with mean score, *Peer assessment* with $\bar{x} = 3,06$ and 37 responses of “Rarely” and “Never”; and *Self-assessment* with a mean of 3,23 and 29 responses of “Rarely” and “Never” were the two least favored methods employed for assessing listening and speaking.

Standardized tests. When standardized test results for listening and speaking assessment were compared, it was seen that the mean score and the percentages of using standardized tests for listening were slightly higher ($\bar{x} = 3,85$) than the percentage of using standardized tests for speaking ($\bar{x} = 3,76$). While, the percentage of “Always” and “Never” responses were similar (Listening: 71% “Always and “Usually”; Speaking: 68% “Always and “Usually”) 20% of instructors stated that they “Never” or “Rarely” use standardized tests for speaking.

2. Findings from the Interviews

In this part, the results of interviews for reading, writing and speaking and listening skills will be presented separately.

a. Methods of reading assessment

According to the interview results, it can be seen that instructors claimed using more than one method during classroom assessment. While the most popular methods were multiple-choice, true/false, matching and student summaries, they indicated that they do not use reading aloud, self-assessment, peer assessment and portfolios. The interview findings also showed that the standardized exams like quizzes, midterms and finals are dominant forms of reading assessment.

Regarding the reasons for using multiple-choice questions, instructors mentioned their practicality, reliability issues, and the school policy while preparing for the exams and the coursebook:

They are all multiple choice because our students are generally used to these kinds of questions, and since high schools, they are not used to open-ended questions in reading texts. (Instructor 1)

Instructor 1 points out the practicality of using multiple-choice questions. As students were more used to being assessed by them, they had difficulty forming open-ended questions.

Instructor 4 also brought up the issue of practicality and reliability for preferring them. According to Instructor 4, the high number of students makes it hard to standardize the students' answers but it is clear from the statement that s/he is aware that they are not the best options.

I believe it is difficult to be standard on the assessment process among teachers because you know, prep departments are crowded and we have so many students, the assessment process should be [reliable]. Teacher A grades differently and teacher B grades so differently. That would cause problems. So that's why, maybe it's not the best way, but this is the most practical way I guess. (Instructor 4)

Instructor 2 gave the reason why multiple-choice questions are not the best way. S/he stressed that although multiple-choice questions are preferred by the institutions as they are practical, they prevent students from developing creative thinking skills as assess lower comprehension skills in an isolated way:

I really don't like them because. You know all that our students cannot think of creative answers and multiple choice does not help with them but our exams mostly are in multiple choice. That's why [I use multiple choice items] just to make them practice. (Instructor 2)

Instructor 3 also supported the belief that it is usually the choice of institutions and instructors feel they have no other option:

We have the exams that the school, you know gives us generally multiple choice. (Instructor 3)

We try to use the questions which are prepared within the course book. (Instructor 6)

Nevertheless, the instructors indicated that they do not depend solely on multiple choice questions and they try to provide students with opportunities to discuss and have a deeper understanding of the texts through brainstorming, summarizing, discussing the multiple choice or true/false items and performance tasks:

But even if it's true or false because it's based on whether it is true and false and if they only have a 50% chance of answering by chance so I always want them to refer to the text or discuss the answers with their friends (Instructor 3)

And of course, when we work on a reading text, first activating this schema, of course, we talk about the topic. We asked them discussion questions, they share their ideas, they try to discuss so we get ready for the topic and then after the reading text again, I asked them some discussion questions to share their opinions (Instructor 4)

When they summarize, I understand that they read the text and then after reading it, I want them to take notes about it. I mean, making shorter notes about that text or the paragraph so I can understand they have an opinion about the text (Instructor 5)

We want their students to practice authentic reading. We want [them] to practice reading by presenting it in groups. We give some criteria. We just want to make them read a book, or a short story

(Instructor 6).

Using follow-up questions, as expressed by Instructor 3 helps understand whether the students answer the questions by chance; using discussion questions as stated by Instructor 4, during pre-while-post reading gives opportunities to students create schemata, react to the texts and personalize the topic by sharing their opinions with other students; summarizing as pointed out by Instructor 5, allows students to identify relevant information and integrate receptive and productive skills; asking students to do presentations in groups based on the authentic readings, described by Instructor 6 let students cooperate and collaborate. It also makes students understand the text in depth.

Lastly, reading aloud was the least favored assessment method among 9 instructors as they do not find it authentic, they do not prefer to assess students' pronunciation in reading and it is not a good method to check comprehension:

I don't use reading aloud. It tells us about the pseudo-performance because in normal life we don't read aloud. It is not authentic, so why should I assess this now? (Instructor 7)

Instructor 7 stated that s/he did not believe the usefulness of reading aloud as she believed that it did not give students a chance to practice in real life. Instructor 5 also agreed with Instructor 7 on the limited usage of reading aloud as it has no value in assessing comprehension.

Reading aloud has nothing to do, just a waste of time in my opinion. I can understand, I can assess them, I mean pronunciation, spelling, everything in other methods in another way, it's not reading aloud. (Instructor 5)

On the other hand, although Instructors 5 and 7 agreed on the point that reading is not a useful method, Instructor 2 had no negative opinion on that reading lessons may serve to practice pronunciation:

When you read something aloud, it doesn't help you to understand a text, it's only a pronunciation activity. (Instructor 2)

b. Methods of writing assessment

According to the interview results, it is revealed that the type of methods

differs according to the level of the students. Instructors pointed out that they use different methods such as short paragraphs (describing a person or opinion paragraphs) or formal or informal e-mails for lower levels, while for higher levels all of them mentioned they use short or long essays. Multiple-choice was the method that none of the instructors preferred.

The majority of the teachers frequently stated that they thought of assessment as a process rather than a way to assign grades. As a result, they all brought up the scaffolding techniques used before writing, such as finding ideas and displaying samples.

The most common theme was scaffolding students before starting writing:

The main problem for our students is generally coming up with ideas so I had to give them a route that they could lead in. (Instructor 1)

Instructor 1 stressed that students find it challenging to find ideas and using scaffolding may help students to generate them before start writing. Instructor 2 also supported this belief and offered a variety of ways such as brainstorming, discussion, debate, or a research task, etc. for students to generate ideas:

I first use that like maybe brainstorming a discussion, a debate, something to generate ideas, to help them generate ideas. Sometimes I give a research task. They do research about it and then share what they find in their groups, and then they outline. (Instructor 2)

Instructor 2 also pointed out the difficulty of expanding ideas in addition to finding ideas and indicated the necessity of spending enough time elaborating on the ideas:

Our students generally have difficulty expanding ideas, even if they know how to write them in an organized way. They don't know how to expand their ideas [so we] spent a lot of time on brainstorming examples, details, you know, expanding those ideas. (Instructor 3)

Another common theme in the instructor's answers was the necessity of providing samples to students at the beginning:

First of all, we need to introduce the type of product we want.

First, we need to work on the samples we need to understand. Let's say for example when we start transforming paragraphs into essays, we show them enough samples for them to understand how things work, how each part works. (Instructor 5)

According to Instructor 4, it is crucial to give samples to show the format and requirements of different writing formats. This way students will have a clear picture of each element of the task. Instructor 5 was also in favor of giving samples to students. S/he believes that thanks to samples, students will understand what is expected from them.

And then maybe a sample can be shown and then we can ask them to write something like that, a paragraph for an essay so they can understand what we expect from them. (Instructor 5)

In addition to giving a sample, students may be given more than one sample. Instructor 6 highlighted that instructors need to provide at least three samples to students. S/he also indicated that coursebooks are good sources for samples and they can be integrated into writing:

Before that, of course, as I said, because we use the course book, there are lots of paragraphs or essays, and maybe during the lesson I just try to show them two or three samples, at least until the, let's say in-class writing time. (Instructor 6)

The variety of methods used to support student learning and enhance writing proficiency—from presenting samples at the start of learning activities to incorporating them into lessons—demonstrates how samples may be used with flexibility.

The effect of standardized exams on the selection of tasks was another theme. 5 out of 10 instructors stated that they used the tasks or methods that were from either booklets or coursebooks and the effect of washback was clear from their responses:

In class writing, I don't grade them but of course, we have writing parts in the exams so we have a criterion and we show the criteria during class writing and explain how they get points according to that criteria in the exam. (Instructor 3)

It is clear that Instructor 3 acknowledged the effect of in-class writing activities on exams, therefore she tries to make students understand how they are graded. Instructor 6 also supported that exams have an effect on classroom assessment and even though she does not favor multiple-choice methods, she revealed that exams are regarded as more important so they cannot escape using multiple-choice methods as they are imposed on them by the administration:

We started using just another effective writing course book or each type of paragraph writing to essay writing, and then we try to do some in-class writing. Actually, this is just for practice. Again, they are just trying to practice writing in the classroom until the real graded one in the class, but they do it twice before the real one. (Instructor 6)

In parallel with this, Instructor 8 highlighted the time constraints and the obligation to prepare students for exams:

I just give feedback to students. I actually take the papers and show every mistake that they made and I wrote the correct sentences by pencil. I want them to understand their mistakes as we have limited time before the exam. (Instructor 8).

c. Methods of listening assessment

The findings from the interview results show that instructors use a diverse range of methods to assess listening. Also, it revealed that instructors use instructor-made methods such as multiple-choice, taking notes and summarizing more commonly than student-conducted methods.

The most frequently used method by instructors was found to be note-taking by 6 instructors mentioning it. The reasons were stated as:

I always try to use Cornell's note-taking method in listening lessons to check their comprehension. (Instructor 1)

Instructor 1 mentioned that s/he uses the Cornell method to check students' comprehension. In Cornell methods, students need to organize the information and this may help them comprehend the listening tasks better.

Instructor 2 described the process of note-taking and emphasized that s/he uses it even though the activity does not directly require it since she believes that it

can contribute to effective listening.

Even if the exercise, doesn't demand any note-taking, I want them to do it so, so I try to use sheets or sometimes based on the principle of the listening extract, I prepare a chart as it has a role in attentive listening. (Instructor 2)

Instructor 5 also supports note-taking for a different reason. For Instructor 2 taking notes is essential, particularly for university-level students because, in their departments, they will be expected to listen to the lectures and take notes. Therefore, she tries to prepare them for their departments.

Another popular method was summarizing but instructors also combine it with the student-conducted method of retelling a story as they want to combine listening and speaking:

Retelling a story... After listening to a passage, they retell a story or the summary of the text, what they listen. It is also a listening and speaking assessment. After listening to a text or video, I want them to summarize the text to each other because I want them to listen to the text carefully. Also want to give them a purpose to listen. (Instructor 2)

For Instructor 2, retelling a story assesses understanding while also encouraging students to actively engage with the content, which develops their attentive listening and effective communication skills.

In addition to retelling a story, Instructor 3 also stresses the role of summarizing in active listening and collaborative learning and she provides some useful ways such as jigsaw listening:

Sometimes I want them to summarize what they have heard and sometimes we do jigsaw listening too, so that they can compare and share their notes. I try to include collaboration in listening. (Instructor 3)

Another example was presented by Instructor 9:

One month ago, there was a listening section about Maradona and God's hands. You know, he was touching the ball before he scored the goal, and I found a video on YouTube. And first, they watch the video, which lasted for one minute. And then I asked them to summarize it to

each other to create a schema. (Instructor 9)

Similar to Instructor 3, Instructor 4 indicated using collaboration by peer assessment by integrating summaries and peer assessment. This way, a supportive environment where students can be involved in the assessment process by giving/receiving feedback from each other is created.

Also, we try to make them interactive when they share their video and the other students should comment on their performance about the topic they talk about or summarize what they do so it would be a good practice, good assessment, peer assessment kind of way for them. (Instructor 4)

Apart from methods, Instructors' responses also hinted that they do not define learning outcomes themselves as they are prepared by the curriculum office so they mostly stick to the syllabus:

In class, we don't give a grade but I look at the curriculum and in our curriculum there are some skills that we need to be covered. (Instructor 3)

Instructor 3 highlighted that the learning outcomes are predetermined by the institution and they had to adhere to them. Similar to Instructor 3, Instructor 7 indicated that the objectives only cover the ones that students are responsible for the exam:

We have our own materials prepared by the institution and we just focus on those objectives to be achieved and to be tested in the exam. (Instructor 7)

Similarly, one instructor mentioned that they have no control over the task types while the other talked about its negative effect on students' motivation:

We only do it multiple choice. Actually, it is not about me. Actually, it is about the Testing Unit. That is why, we actually prepare materials according to them. (Instructor 8)

A lack of flexibility over task kinds is highlighted by Instructor 8, who mostly uses multiple-choice questions that are set by testing units. Instructor 9 also mentioned that when there is a discrepancy between curriculum and assessment, it

has a direct effect on students' motivation to participate in lessons:

We don't assess their speaking and listening abilities and that's why they don't value as speaking and listening activities in the book so they are not very eager to participate in those activities. (instructor 9)

While some instructors accepted it, some other instructors made changes to the assessment methods:

Also, I asked some questions that are not in the book. Although in the exam, there are multiple choice questions, I also ask them detailed questions, like critical thinking questions. (Instructor 1)

By introducing methods such as critical thinking questions, it is notable that Instructor 1 noticed the lack of methods in standardized exams for assessing high-order skills such as critical thinking, and this way s/he may be trying to close the gap and expand the opportunities for students and foster higher-order thinking skills. Similarly, Instructor 6 tries to create chances for students to communicate and improve assessment methods beyond the standardized tests.

But if it is based on the course book, mostly they are exposed to multiple choice questions but I also try to let's say make them speak while listening so after listening, I also have some open-ended questions for them. Maybe they have a chance to show what they understand by speaking as well. (Instructor 6)

d. Methods of speaking assessment

Unlike listening, interview findings showed that the instructors use more student-conducted methods to assess speaking. The most popular methods were oral presentations, oral interviews and oral discussions. In addition to that, all instructors emphasized that they do not use only one method in the classroom.

Some instructors pointed out the benefits of using presentations as:

Students are reluctant to do anything because they are afraid of public speaking so I generally encourage my students to have presentations in the 3rd or 4th week as I think they will have more confidence when they speak in front of the examiners. (Instructor 1)

Instructor 1 highlights the value of presentations in helping students get over the anxiety of performing in front of others by making them practice in a supportive environment rather than expecting them to speak in front of the examiners.

Another benefit of the presentation was indicated by Instructor 10:

We use presentations as they are going to do presentations after they graduate. It is necessary for them to learn how to do a good presentation (Instructor 10).

Instructor 10 highlighted how important it is to make students ready for professional environments, where good presentation skills are frequently needed. While both teachers encourage the use of presentations, Instructor 10 places more emphasis on getting students ready for the demands of the workforce, while Instructor 1 places more emphasis on helping students overcome their fear of public speaking and develop confidence. Different priorities when using presentations as a teaching tool are reflected in this disparity in focus.

Secondly, instructors also mentioned the level of the students are also determining the task type or task difficulty:

For lower level sometimes, picture description as well. (Instructor 7)

But for the first term, we used again a recording I guess for elementary and pre-intermediate level, but it was about a description of a photo. They showed the photo and they tried to describe it, maybe according to their last summer holiday or their favourites film. (Instructor 6)

We try to prepare the questions based on the topics that we covered in the classroom and. For lower levels, we have more daily use. (Instructor 3).

Overall, all three instructors highlighted the necessity of tailoring the task types according to students' level and they indicated it is especially needed for lower levels.

Another common theme mentioned by two instructors was the problem of reading:

I don't like dialogues because if we give them some samples, they memorize them and then come back I like oral discussions more when I use them in their original environment, not like the speaking exam but while they are talking to me or giving an answer or talking to each other in English. (Instructor 4)

It is clear from the statement that Instructor 4 had a concern regarding the memorization of dialogues therefore s/he places more emphasis on oral discussions probably as they foster more natural communication. Instructor 9 gave information about how to utilize discussion in the classroom by giving a time limit:

I don't want them to read what they wrote. I urge them to speak. I want them to find a partner and discuss the questions. I give them a time limit and then we have a whole class discussion. (Instructor 9)

Another theme occurring during interviews was providing enough input. One instructor highlighted the importance of providing samples or enough input for lower-level students to make them feel ready to speak:

We prepare lower-level students with some, let's say, samples through thematic questions taken from each unit based on the first book and in the classroom. Each week, actually, we teach them some speaking strategies, like one week, let's say turn-taking strategies, another one agreeing, disagreeing. Without giving this input, we never expect those students to perform really well because they need this input first. (Instructor 7)

Instructor 7 described how they prepare students to speak confidently step by step and stressed that students need to be given the necessary input and be taught some strategies such as turn-taking, agreeing, or disagreeing. This may be a solution to students' memorization as they will feel more confident.

The last theme according to the interview results was the assessment of standardized exams. While half of the instructors indicated that they assess students' recorded performances or presentations themselves, the other half said they have a partner and both of them give grades by showing them to each other.

Sometimes we watch videos on our own and evaluate those based on the criteria again, but standardization sessions are generally held

before the speaking exam, and they provide us with two or three video recordings from the previous years, and we come together as groups in the meeting and again, we give our schools in. In the end, we all discuss and come to or try to reach a consensus on the point. (Instructor 7)

Instructor 7 explains how students' performance on standardized tests is assessed with one instructor. It is clear that Instructor 7 was aware of the reliability issue so s/he explained that to guarantee reliability in grading between assessors they held standardization meetings before. On the other hand, Instructor 10 mentioned there is another way of ensuring fairness and this is to use two instructors during the assessment:

During interviews, one teacher asks the question and the other teacher listens. Then they give their marks according to the criteria without showing to each other. Then they come together and discuss it if there is a discrepancy. [I believe] this is to be objective. (Instructor 10)

Although both instructors use criteria for ensuring reliability, unlike Instructor 1, Instructor 10 uses a more reliable method by discussing the score with another instructor before reaching the final decision.

C. Findings Regarding the Beliefs of EFL Instructors

The third research question was about the beliefs of EFL Instructors on Classroom assessment. This part focuses on the questionnaire and interview results.

1. Findings from the Questionnaire

In order to examine the beliefs of EFL Instructors about language assessment, the instructors were asked to respond to 8 items. While the first item is related to beliefs about summative assessment, the other 7 items are about formative assessment. Table 8 reflects the beliefs of EFL Instructors about language assessment. The data are sorted in descending order, from the highest mean to the lowest.

The data analyses reveal that the majority of the instructors believed that feedback has a positive effect on students' improvement. First of all, 97% of

instructors either strongly agreed or agreed on the necessity of giving feedback to students so that they understand how to improve their work and progress in their learning. Similarly, regarding peer feedback more than half of instructors 59% strongly agreed that instructors should provide opportunities for peer feedback and self-assessment to enable them to reflect on their learning. In addition to that, 89% of instructors either strongly agreed or agreed that when the students are involved in self- and peer assessment can make them more motivated and improve their success. There were only 2 respondents who disagreed with the statement. Table 9. presents the interview results regarding the frequency of self and peer assessment usage among the instructors based on the data retrieved from the methods of reading, writing listening and speaking assessment.

Table 8. Beliefs about Language Assessment

Principles	Mean	Median	Mode	SD	SA N %	A N %	NAND N%	D N (%)	SD N (%)
4-Students should be given feedback which enables them to know how to improve their work and take their learning forward.	4,74	5,00	5	,543	78	19	2	1	0
6-Instructors should first identify the purpose of the assessment when they design the assessment task/test.	4,67	5,00	5	,620	74	20	5	1	0
3-Students should be given opportunities to reflect on their learning through peer feedback and self-assessment.	4,51	5,00	5	,659	59	34	6	1	0
8-Involving students in self- and peer assessment activities can enhance their achievement and motivation	4,41	5,00	5	,740	54	35	9	2	0
7-Students will learn the most and perform their best if they know in advance the criteria that will be applied in judging performance.	4,38	5,00	5	,776	53	35	9	3	0
2- Assessment results should be used to modify instruction to Better support students' learning.	4,32	4,00	4	,665	41	52	5	2	0
5-Instructors should negotiate the assessment criteria with their students in advance.	4,18	4,50	5	1,019	50	29	11	9	1
1-Assessment should be always associated with marks	2,52	2,00	2	1,000	5	12	23	50	10

*SA=Strongly Agree; A= Agree; NAND= Neither Agree nor Disagree; D= Disagree; SD= Stringly Disagree; *SD = Standard Deviation

The data analyses reveal that the majority of the instructors believed that feedback has a positive effect on students' improvement. First of all, 97% of instructors either strongly agreed or agreed on the necessity of giving feedback to students so that they understand how to improve their work and progress in their learning. Similarly, regarding peer feedback more than half of instructors 59% strongly agreed that instructors should provide opportunities for peer feedback and self-assessment to enable them to reflect on their learning. In addition to that, 89% of instructors either strongly agreed or agreed that when the students are involved in self- and peer assessment can make them more motivated and improve their success. There were only 2 respondents who disagreed with the statement. Table 9. presents the interview results regarding the frequency of self and peer assessment usage among the instructors based on the data retrieved from the methods of reading, writing listening and speaking assessment.

Table 9. Using Peer and Self-assessment

Method	Frequency N (%)				
	Always	Usually	Often	Rarely	Never
Peer assessment					
Reading	13	26	39	15	7
Writing	16	28	24	28	4
Listening & speaking	16	25	22	23	14
Self-assessment					
Reading	16	23	42	15	4
Writing	17	33	27	18	5
Listening & speaking	19	25	27	18	11

According to Table 9, instructors' use of peer assessment was more frequent for reading than for writing and listening & speaking. Overall, 39 instructors stated that they "Usually" or "Often" use peer assessment for reading. This result may indicate that although 93% percentages of instructors believe that instructors should provide opportunities for students to reflect on their teaching through peer feedback and self-assessment, only 78 % of them "Always", "Usually" or "Often" indicated that they used peer feedback for reading and 68% percentage of them "Always", "Usually" or "Often" use it for writing, and 63% of them stated that they "Always",

“Usually” or “Often” used it for listening & speaking. Concerning self-assessment, the data indicated that the instructors use self-assessment for reading more than they use peer assessment for reading. Also, the overall percentages of the writing and listening & speaking for self-assessment were higher (71%) than the overall percentages of peer assessment (63%). In total among the 93% of instructors who agreed or strongly agreed on the necessity of peer assessment and self-assessment, 81% of them stated that they use self- assessment for reading, 77% of them use it for writing and 71% of them use it for listening & speaking.

The second item with the highest mean score ($\bar{x}=4.67$) was about identifying the purpose of the assessment before the tasks/tests. 74% of instructors strongly agreed that teachers should determine the criteria before creating the assessment tasks/tests. In parallel with this, more than half of instructors (88%) either strongly agreed or agreed that if students are aware of the criteria that will be used to evaluate a performance before, their learning and performance will be maximized. While 50 out of 100 instructors strongly agreed that assessment criteria should be determined by negotiating with students, 11 instructors stayed neutral, 9 instructors disagreed and 1 instructor strongly disagreed. Item 1 was about assessment of learning and was related to summative purposes. Most instructors either disagreed 50% or strongly disagreed 10% that assessment should always be linked to grades, but 23% of instructors stayed neutral. Except for the first one, all other items in the table investigated instructors’ beliefs towards formative assessment, AfL and AaL. Related to using assessment results to scaffold learning by adjusting their teaching, 93% of instructors either “Strongly Agree” (41%) or “Agree” (52%) options. As it is referred from Table 8, nearly all instructors have positive beliefs towards using classroom assessment for learning.

In addition to beliefs, the final part of the questionnaire was about the type of feedback instructors provide to their students. The type of feedback is shown in Figure 5.

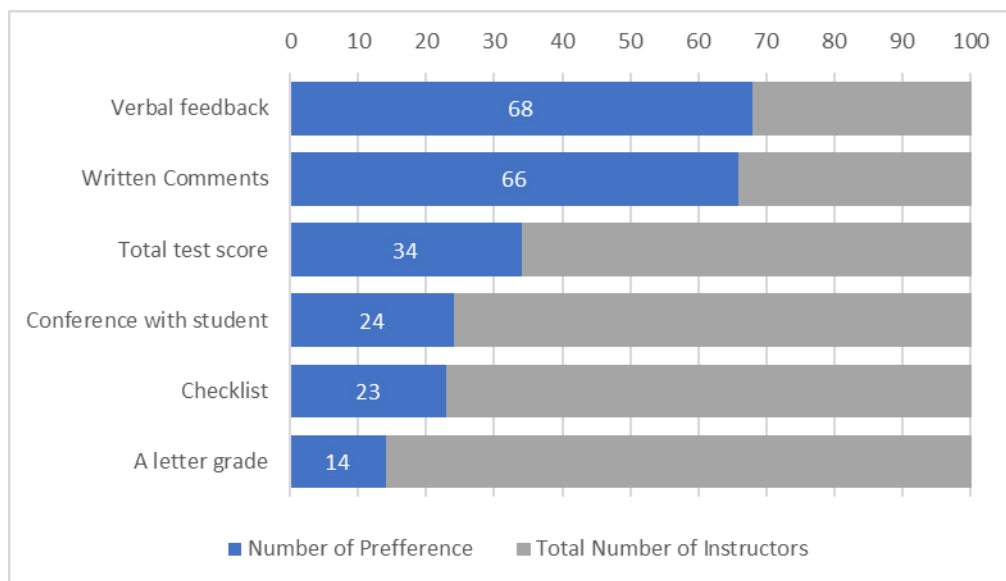


Figure 5. Types of feedback provided by instructors

Instructors were asked to state the preferred type of feedback in the questionnaire and select the type(s) of feedback they give either in or outside the class. Figure 5. shows that the main form of feedback was verbal feedback (N=68). Following this, written comments were the second most preferred feedback type (N=66). The number of instructors preferring total test scores (N=34) and total conferences with student (N=24) were nearly equal. While checklist (N=23) was the fifth most popular feedback type among instructors, very few instructors preferred a letter grade (N=14).

2. Findings from the Interviews

The interview data revealed valuable data regarding the instructors' beliefs on giving feedback, the effect of peer and self-assessment on students' learning and motivation, providing criteria and the way of using assessment results to enhance learning and teaching.

Feedback: First of all, the themes according to the data included the type, manner and time of feedback. While the most frequent methods for giving feedback were mentioned as written and oral feedback, two of the instructors also mentioned they use conferencing when students demand it. First of all, most of the instructors mentioned giving timely feedback is vital for students to progress. They indicated the feedback can be instant or process-based according to the type of activity or skill. During in-class activities of reading, listening and speaking, instant feedback

focusing on the error was common:

That's based on the type of activity like fluency or accuracy. But if it is not an accuracy activity, then I don't actually. I don't stop them, because they're already shy to speak, so if I, you know, stop them from every mistake that they make, then that will be a very dull activity but if it's an in-class activity and if the activity needs instant feedback, I move around the classroom and I'm monitoring the students. I give instant feedback. (Instructor 3)

Instructor 3 indicated that s/he modifies feedback according to the type of activity. As interrupting students will cause anxiety or refrain them from speaking at all, s/he suggested that instructors should not focus on every single mistake. Similarly, instructor 8 indicated that s/he uses monitoring:

For listening and speaking activities, I ask them to find a partner, and then discuss the question. I put a time limit for this, like 5 minutes or 10 minutes and they discuss the questions with their partners. I walk around them and observe them and give them again instant feedback if they have any questions. (Instructor 8)

Like Instructor 3, Instructor 8 stated that she employs monitoring and assistance if students demand it. Although it is not clear how Instructor 3 and 8 gave feedback, Instructor 9 gave a detailed description:

The main reason is to provide feedback, provide my students with feedback, and the feedback is like an instant one. I give them instant feedback. It makes them like learn so they progress more smoothly. For instance, to check their pronunciations, to see if they make mistakes. When they finish speaking, I just repeat the same sentence or I read it a lot too, so I make them try to understand their mistakes and I make them learn the pronunciation correctly. (Instructor 9)

It is clear from the statement that, like the other two instructors, Instructor 9 also acknowledged the role of instant feedback on student learning. While Instructor 3 seems to directly give feedback by telling them, Instructor 10 seems to answer students' questions. On the other hand, Instructor 9 uses prompting so that students will discover their own mistakes in the areas of accuracy and pronunciation.

Regarding the manner of feedback, 5 instructors highlighted that the feedback should not only focus on the weaknesses of the students. It should also include positive sides to motivate them:

Maybe after the activity or maybe at the end of the lesson, I write the mistakes on the board. I never say that this is this person's mistake. For example, today I gave feedback. Yesterday, they wrote essays and today I took some good sentences or problematic sentences from those essays. And dropped them on a worksheet and ask them to correct the mistakes, find them, correct the mistakes. They didn't even know who wrote these sentences. I said so that somehow, I gave some feedback. The sentences were chosen on purpose. Actually, there were some common mistakes and a good sample of language. That's why I chose them. (Instructor 1).

While sometimes I give the whole class feedback on the common mistakes, for the lower levels to praise them and to make them work on their skills better, I give individual feedback. (Instructor 2)

I mean while giving feedback maybe using the sandwich methods and the first the good parts of the students and then some negative and then again the positive ones can influence in a positive way. I tried to give positive feedback during the lessons. OK, this motivates students a lot, but if it is something negative, they should change something. I mean, I do it after the lesson privately. (Instructor 5)

For speaking only I'm doing it verbally by analysing the comments they made. And I say if you still would like to talk about your mistakes 1 by 1. You can come and visit me in break time. Then we can maybe have a look at it. Yeah, because sometimes they don't feel, I don't know, comfortable if I just call their names in the classroom. (Instructor 6)

But I try not to do that in front of all the students. I tend to be private, by using conferencing, outside class hours in order not to discourage them. I usually include the weak parts. Because they want, they require more analysis, more detail. But of course, I make a positive

comment, a generally positive comment based on the strengths of the paper. (Instructor 10)

According to the interview data, it is clear that all instructors share the common belief that students need constructive feedback, but their methods change according to their style. While the whole class feedback, as expressed by Instructors 1 and 2, helps instructors address common mistakes, giving negative feedback in private, described by Instructors 5,6 and 10 avoid students being uncomfortable and discouraged. For private feedback, either break times or out-of-class are proposed as examples. In addition to giving feedback on weaknesses, feedback on students' strengths as stated by Instructors 1, 2, 5 and 10, is presented as a tool for motivating students to progress especially for lower-level students.

Although the instructors gave students one-shot feedback during the activities, the interview data also showed that according to the type of skill, the feedback can be process-based. 7 out of 10 instructors indicated the importance of reflection:

Students know they have different learning styles and also, they need, helpful feedback during the process. (Instructor 4)

Instructor 4 highlights the necessity of tailoring the feedback according to different learning styles. That brings us to the point that instructors need to combine different feedback types together since whole class feedback is not suitable for students with different needs. Furthermore, while addressing the problematic areas directly can make students feel safe, the necessity of reflection is also stressed by Instructor 5:

We always provide them with some sheets to put in the folder. They keep that always and they see my written feedback and my oral feedback, both of them and also for writing, we show them the common mistakes. They also as a second step, I mean reflect on their own mistakes. like, some plans for a feature like, you know, how to improve these mistakes (Instructor 5)

Giving students their tasks with comments on them may make it easier for students to go back and refer to those again when they make a similar mistake again giving students a chance to reflect on their mistakes and make future plans to correct

them. In addition to using comments, correction symbols as stated by Instructor 6, serve the same purpose:

We have just the symbols and we can also have a little feedback in a written version of their papers, and then they try to understand their mistakes. Of course, they can't understand the symbols let's say at the very beginning of our feedback. But then we also add detailed feedback in a written version like a little short paragraph. And then they understand their mistakes. Then they correct it. So, we use both symbols and also the feedback. (Instructor 6)

Although getting used to symbols may be challenging for students at first, Instructor 6 emphasizes that they also combine it with detailed feedback till they get used to it. The rationale behind it is to promote self-autonomy for learners.

Self and Peer-Assessment: Seven instructors emphasized the inclusion of students in the assessment process with self and peer-assessment especially for speaking and writing:

Again, we assess their performance in class so that they can see what they lack and what they're good at so that they can see each other's product and comment on each other's product. I think we have to do peer and self-assessment because they need to understand what they're doing. I generally tell them if it is writing as I, especially the other teachers as well of course tell them to keep all their products so that they can see their improvement like and when they are in a second term. I ask them to check your first piece of writing and now look where you are. Now you're able to write a complete essay. It makes them motivated (Instructor 4)

Instructor 4 pointed out that self and peer assessments are useful tools in order to help students identify their strengths and areas of improvement. Through these activities, students may gain a deeper understanding of their learning process and the tasks they are working on. Keeping the products and seeing their improvement over time make the learning visible for students:

I frequently use peer assessment because, in the peer assessment, they are both students and teachers. While they are checking their friends'

writing tasks or anything, I mean essays or paragraphs, they can also understand their mistakes, so I like that peer assessment and I use it a lot. If you monitor and correctly give feedback, in the correct method, they do it perfectly. (Instructor 5)

Like Instructor 4, Instructor 5 also emphasized that self and peer assessment have dual benefits for students since they learn from their own mistakes while interacting with peers. However, while Instructor 4 focused on the benefit of self and peer assessment for students, Instructor 5 stressed the importance of careful monitoring while students are assessing each other's work. While the two examples are about using self and peer assessment for writing, instructor 7 pointed out that it can also be extended for speaking assessment:

We provide our students with sheets feedback sheets whenever we carry on and speaking activity or writing activity in the classroom, we expect them to fill in the sheets. We group them and we allow two or three groups during the exam time and the other groups don't just sit and wait, they need to give feedback to each other. They are not graded. They just need to give their comments to each other. (Instructor 7)

Unlike the other two examples, Instructor 7's approach suggests integrating oral interaction into assessment. It is different from the other two examples in the way that, students work in groups, and assessment is utilized not in class time but during the formal exam.

Although most instructors use peer-assessment in the classroom during writing, 3 of them mentioned the difficulty of applying peer-assessment due to time constraints and the proficiency of students:

Because of lack of time, I don't prefer using peer assessment, even if I know that I have to. If you are working with lower levels, sometimes peer assessment is a waste of time because they don't know how to give feedback. [Also], I think they cannot edit sentences or make complex sentences because they also lack some grammar knowledge. (Instructor 3)

In discussions, they try to assess themselves but normally because of time limitations, I just correct their mistakes at home. (Instructor 8)

To be honest, I rarely use peer assessment because I don't think they're competent enough to assess each other's work. (Instructor 10)

The alleged challenges, expressed by Instructor 3, and Instructor 8 may shed light on the reluctance of instructors to use peer assessment due to feasibility. Instructor 3 and Instructor 10 expressed his/her concerns regarding the students' lack of proficiency and skills. In a similar way, Instructor 8 stressed the time spared for classroom activities is limited and most instructors have to ensure they cover some topics, which forces them to spend less time on peer assessment or skip it completely.

Assessment Criteria: Regarding informing students about the assessment criteria before the task, all the instructors indicated that sharing criteria helps students understand what is expected from them. They also mentioned that in addition to sharing the criteria, instructors need to clarify the criteria by providing samples and examples:

In the first week, we discuss assessment criteria to clarify, to make sure everything works best with each item. (Instructor 1)

The necessity of sharing the criteria in the very beginning, expressed by Instructor 1 shows that students need to be well informed about what is expected from them and they have enough time to understand each point. Instructor 2 also highlighted the importance of clarification of items in the criteria:

I shared the assessment criteria the school sends then we looked at the items in it and talked about what they mean. We talk about the examples. (Instructor 2)

I even, you know, give weak and strong papers, for example, in the classroom, I sometimes write them. Sometimes I use one of the things, you know, that is in my folder. I tried to explain the criteria one by one. I think they should know what we expect from them. (Instructor 3)

The role of discussing each item with students on clarification was further elaborated by Instructor 3. In order to clarify points, Instructor 3 indicated that students are given both strong and weak examples so that students will be able to see the criteria with concrete examples. Instructor 7's comment implies that the criteria

should be based on learning outcomes and in a way align with the curriculum to be valid:

Also, we know they know which objective they are responsible for. So as syllables, they have separate outcomes and objectives for each type of lesson skill, so they know. (Instructor 7).

Besides, the interview results also showed that 6 of the instructors used assessment criteria for preparing students for the exams. Some samples can be seen in the following comments:

Otherwise, students really wonder and ask “Teacher, by which criteria did you just evaluate our scores?”. (Instructor 6)

Instructor 6 underlined that students want to know the criteria they are assessed for so instructors need to be transparent to students in order to prevent confusion. In order to prevent confusion or unexpected results, Instructor 3 also asked students to grade themselves according to the criteria:

I even asked them to grade. So, what would you give in the task achievement part Do you think it's worth eight points or 4 points etc. And when they do that, I think after the exam, if they get a lower point than they expect, they won't be surprised. (Instructor 3)

By engaging students in the self-assessment process and asking them questions, Instructor 3 shows a model for preparing students to take responsibility for their own work. While exams may be used for formative assessment purposes like this, sometimes the effect of exams can be debilitating for learning:

In writing lessons, they know which criteria they will be assessed, so I give them criteria, but I don't give them any criteria before speaking because they don't have any speaking exams. (Instructor 9)

Although Instructor 9 stated that for writing exams criteria are given to students, when the students are assessed in the classroom in an informal way, the instructor does not give them any criteria which can make speaking challenging for students.

Assessment Results: The interview data showed that 8 instructors use assessment results mainly for formative purposes. The most common team was

modifying their own instruction:

After the exam, if I notice it, I change my instructions. I also change the topics that I'm going to cover. I may go back to the objectives that I have already covered. (Instructor 1)

Instructor 1 placed a strong emphasis on adopting a student-centered approach which requires going over previously practiced topics and revisiting the learning objectives in case of needs. Similarly, Instructor 2 indicated that exams are also beneficial for tracking students' learning:

If they make lots of mistakes in the exams, I try to spend a lot more time on this area by keeping their weaknesses in mind and I take them into consideration and act according to that in my lessons. (Instructor 9)

While Instructor 1 stressed changing the topics, Instructor 9 adopted a more general approach and spent a lot of time on working problematic areas to ensure meeting special needs. Instructor 6 goes further by stating that s/he collaborates with other colleagues to diagnose their own weaknesses and curriculum-related problems:

I mean, we also tried to share my students' assessment results with other instructors in a meeting because we, I mean, if you find any score negatively, we try to find our mistakes, so we need to change our teaching or modify the curriculum. (Instructor 6)

With this approach, Instructor 6 emphasized that in addition to remedial teaching, it is necessary to discuss instructional or curriculum-related issues collaboratively to achieve quality assessment.

While the majority of instructors believe that modifying the instruction is crucial for improving students' learning, two instructors indicated that they do not modify their instruction due to the hectic schedule or curriculum-related issues:

After the exam, we check the sections of the students' knowledge according to that we prepare ourselves for the next module to help them better in a positive way. Umm, so you cannot do it while teaching your lesson because like it already passed you, you don't have any effect on the curriculum yourself. (Instructor 8)

A characteristic of summative assessment is assessing students at the end of a specific period of instruction and the results are usually presented in grades. Therefore, it does not provide students with opportunities to revise their work. Likewise, it can be understood from Instructor 8's words that instructors do not have opportunities to make immediate adjustments to teaching. On the other hand, Instructor 8 emphasized that the assessment results can guide the planning process for future learning. Instructor 8 also underlined that they have no control over the curriculum. Instructor 10 also brought up another institution-related problem that hinders instructors from making real-time adjustments to their teaching:

It's really impossible for me to change the lesson because the teaching schedule is too hectic. (Instructor 10).

D. Summary of the Results

1. Summary of the Questionnaire Results

The findings related to the purposes of the classroom assessment showed that instructors utilized classroom assessment for different purposes. The overall results show that instructors use classroom assessment primarily for student-centered purposes to foster students' learning and enable them to control their own learning rather than for instructional purposes. While the weighted average mean score for student-centered purposes was $\bar{x} = 4.06$, it was $\bar{x} = 3.58$ for instructional purposes. The top-ranked classroom assessment purpose among instructors was *"to provide feedback to my students as they progress through the course"* with $\bar{x} = 4.47$. Instructors also stated that they assess students *"to diagnose their strengths and weaknesses"* as well as *"to obtain information on their progress"*. Regarding the instructional purposes, *"to plan my instruction"* was the highest ranked purpose with a $\bar{x} = 3.82$. The second highest ranked purpose was to *"to diagnose strengths and weaknesses in teaching"* with a $\bar{x} = 3.80$. The data also reveal that *"to provide information to the school administration"* was frequently employed purpose among instructors with a $\bar{x} = 3.76$. %37 of instructors stated that they always use it. The least preferred purpose of classroom assessment was stated to be *"to group my students for instructional purposes in my class"* with a $\bar{x} = 3.13$.

To understand the practices of classroom assessment, instructors were asked

to respond to the questions about the methods of assessment for reading, writing, listening and speaking skills. For reading, generally, the results indicated that instructors use both instructor-made methods and standardized tests more than they use student-conducted methods. The most popular methods of assessing reading were instructor-conducted methods of multiple choice, ($\bar{x}=3,91$), true-false ($\bar{x}=3,82$) and matching items ($\bar{x}=3,82$). Among the student-conducted methods, the highest mean scores belong to student summaries of what is read ($\bar{x}=3,41$) while the lowest mean score among all the methods belongs to the student-conducted read-aloud method ($\bar{x}=2,83$).

Instructors' preference for standardized tests was also high ($\bar{x}=3,68$). Regarding writing, essay-style assignments were typically preferred by instructors, and short essays ($\bar{x}=4,12$), were the most favorable. This was followed by editing a sentence / a paragraph ($\bar{x}=3,85$) and long essay ($\bar{x}=3,66$). On the other hand, student-conducted methods of student journals were the lowest-ranked item among all others ($\bar{x}=2,54$). Furthermore, instructors' attitudes on true/false and matching items are divided and are not generally positive. The results showed that instructors also prefer standardized tests ($\bar{x}=3,69$). Concerning listening and speaking, the data revealed that instructors mostly applied student-conducted methods of oral interviews= dialogues ($\bar{x}=4,02$). and oral discussion with each student($\bar{x}=4,00$). Following these, the highest-rated method was instructor-made multiple-choice items for listening ($\bar{x}=3,99$). Among all the methods, peer assessment was the lowest-rated one with ($\bar{x}=3,06$). The data for all assessment methods indicated that except for listening and speaking, instructors tend to use instructor-made methods more than student-conducted methods. Furthermore, instructors use self-assessment and peer assessment more for listening and speaking than they do for reading and writing. Moreover, the mean scores of standardized tests for assessment were quite high for all levels which indicates that instructors employ them as the main part of classroom assessment.

The results relating to the beliefs on assessment, nearly all instructors have positive beliefs towards using classroom assessment for learning as the mean scores were higher than 4.00 and instructors mostly strongly agreed or agreed with all the statements.

2. Summary of the Interview Results

As understood from instructors' comments, the instructors use classroom assessment both for formative and summative purposes. The most emphasized purpose was to obtain information about the student's progress. Within this scope, the instructors stated that they use classroom assessment to check if the classroom objectives were met or to diagnose students' strengths and areas for growth for future learning. They pointed out that in this way, they try to enhance students' learning. To contribute to students' learning, the instructors listed providing feedback as the second main purpose and it is highlighted that feedback should be given throughout the learning process to help them overcome the difficulties of learning rather than at the end of learning. While the instructors also mentioned that they use classroom assessment for summative purposes such as grouping them according to their level during classroom activities or placing them at the correct level, the majority of the instructors stressed that the main aim is to motivate students as well as encouraging them to be involved in the learning process.

According to the interview results, it is found that the instructors rely more on instructor-made assessment types for reading, writing and listening, they use student-conducted assessments for speaking assessment. For reading assessment, the leading assessment types were mentioned as multiple-choice, true/false, matching and student summaries. On the other hand, no instructor was inclined to use reading aloud as it was not authentic and they do not expect students to do pronunciation practice during reading. Although instructors were aware that multiple-choice questions were not the best assessment type as they do not promote critical thinking and creativity, they used them because of reliability, practicality, and school policy. Bearing this in mind, the instructors pointed out that they usually combine them with open-ended questions. To make students understand the texts better, the instructors used summarizing as a supplementing technique.

For the writing assessment, the instructors stated that the students' levels had a powerful effect on the selection of task types. For lower levels, the most common tasks were short paragraphs (describing a person or opinion paragraphs) or formal or informal e-mails, while for higher levels, short or long essays were the leading task types. Also, the dominant effect of exams was reflected on the selection of tasks. The instructors called attention to the fact that they only practiced the task types that

students were tested in the exams. The interview results also showed that instructors scaffold their students during pre- while -post writing by helping them generate ideas, providing samples, and including them in the assessment and feedback process through self and peer assessment. On the other hand, some of the instructors mentioned the difficulty of peer feedback because of time constraints and the proficiency of students.

The assessment methods for listening were similar to reading and one instructor stated that as they are both receptive skills, they have similar practices. Therefore, the prevalent methods for assessing listening were listed as multiple-choice, taking notes, summarizing and filling in the blank. Firstly, although multiple-choice questions were predominant in the coursebook, some of the instructors indicated that they do not limit themselves to them and try to add more methods such as note-taking and summarizing. The instructors asserted that thanks to note-taking and summarizing, students can focus more on what they are listening to. They also underlined that they try to integrate some pair-work or group-work speaking activities into listening. One issue regarding the listening assessment was negative washback. As mentioned by one of the instructors, students were not willing to participate in class activities because they did not officially test listening.

The instructors indicated that they use eclectic methods for speaking assessment. Again, the level of students was determinant in choosing the task types. While for lower levels, picture description or class discussion were used heavily adopted, for higher levels, presentations and oral discussion were used by the instructors more as they help students acquire academic skills. Similar to writing assessments, the instructors mentioned that they provide students with samples or some speaking strategies and do a lot of step-by-step practice before the actual tasks. During the interviews, the instructors implied that students were anxious and reluctant to speak; hence they try to memorize or read their notes as they do not feel ready to speak. In addition to these tasks, summative assessment tasks such as interviews with instructors and were also heavily used by the instructors.

Lastly, the interview data revealed some insights into the instructor's beliefs on classroom assessment mainly about feedback, peer-self assessment, criteria and the purpose of assessment results. The instructors mentioned that they mostly give oral and written feedback and highlighted that feedback needs to be timely and

constructive. Related to self and peer-assessment, while the instructors expressed that they use them, especially for writing and speaking, some of them pointed out the challenging nature of them especially due to students' proficiency levels and time constraints. Regarding the assessment criteria, there was a strong consensus among the instructors for sharing the criteria and clarifying it well with samples and examples. Lastly, according to the interview data, the instructors mostly mentioned formative purposes such as diagnosing weaknesses and modifying their lessons accordingly. On the other hand, two instructors emphasized that although they are aware that showing grades are not enough to enhance learning, they may not be able to do that due to tight schedules and lack of authority in changing curriculum.



V. DISCUSSION AND CONCLUSION

A. Introduction

The present chapter is devoted to the discussion of research questions. While the first three research questions deal with EFL instructors' purposes and methods of classroom assessment, as well as their beliefs on classroom assessment in general, the last one aims to investigate whether the beliefs of the instructors are in line with their practices. For this purpose, the findings of the quantitative and qualitative data will be discussed in the light of relevant studies. The last section of the chapter will provide some suggestions for future studies.

B. RQ1: For Which Purposes do EFL Instructors at Turkish Universities Use Classroom Assessment?

The first research question was about purposes of EFL teachers in classroom assessment. The quantitative data showed that the EFL instructors mainly use classroom assessment for student-centered purposes to provide feedback, to diagnose what students can do and what they need further assistance and to track students' progress. The interview data also supported that the instructors gave considerable importance to the formative purpose of assessment, especially to enhance students' learning and assist them with feedback. In Jannati's study (2005), the majority of teachers (10 out of 18) stated that they used assessment results just to diagnose the strengths and weaknesses of the students and provide feedback. The summative purpose of classroom assessment was also indicated in the interview to be an important purpose to enable them to group their students according to their levels, to encourage them to motivate their students or to show them what can they achieve at the end of the course, the instructors gave greater focus on AoL. It is notable that only one instructor mentioned the purpose of assessment is to test students' knowledge and the best way is the summative forms such as quizzes, midterms etc, while others strongly underlined they either do not prefer to grade students or they do

it when they have to. When viewed from this aspect, AfL seems to constitute part of classroom assessment for these instructors. This finding is in line with the result of earlier research conducted by Önalın and Karagöl (2018) which demonstrated that 70 Turkish EFL teachers who work at the Prep Class of a state university strongly believe that assessments should be used for formative reasons.

On the other hand, although the instructors' willingness to use AfL was obvious from the questionnaire and interview results, it is notable to see that they practice it at a moderate level in terms of teachers' reflection. In the interviews, those who mentioned that they use assessment to diagnose their own strengths or weaknesses were scarce. This result is similar to the findings of Erdol and Yıldızlı's (2018) study which found that while the majority of teachers use classroom assessment to provide feedback, increase learning and determine the level of achievement, very few teachers utilize it to see their own strengths and weaknesses whereas it contradicts to the result of Widiastuti (2020) study. As Hargreaves (2005) points out, the idea of assessment for learning is linked to the presumption that both teachers and students need to understand the direction in which learning is going. Thus, AfL serves as a tool for determining the next action to be taken. This implies that assessment for these instructors is a process of evaluation of student's learning rather than a mutual cooperation to improve teaching and learning. One possible reason for this may be teachers' lack of assessment literacy. Teachers may lack sufficient training on how to use assessment to reflect on their own teaching. Furthermore, another reason can be too much focus on students' learning. Rather than emphasizing the improvement of teaching practices, professional development frequently concentrates on using assessment to make decisions on students' performances. A study by Ashraf and Zolfaghari (2018) supports this claim. According to the study, when instructors' assessment literacy increases, they tend to reflect more on their instruction. In line with this study, the findings from Tajik and Ranjbar's (2017) study showed that limited class time and the pressure to complete a certain curriculum during a semester hinder teachers from having the opportunity to reflect on their lessons and make necessary changes.

C. RQ2: Which Methods do EFL Instructors at Turkish Universities Use to Assess Four Skills (Reading, Writing, Listening, and Speaking)?

Reading Assessment: The findings from the questionnaire and interview revealed that Turkish ELF instructors mostly use instructor-made assessment methods to assess reading in the classroom, yet they also utilize standardized tests and summarizing. This result conforms to Klingner's (2014) claim that teachers need to combine multiple methods while assessing reading. The most frequently used instructor-made methods were indicated as multiple-choice, true/false and matching items. The factors influencing the instructors' choices may be the school's policy and the quality of materials. When standardized tests include those items, it may affect the curriculum and class materials. The negative washback effect is also apparent from the interview results. As the instructors seemed to be aware that their students would be tested with this method in standardized exams, they included those methods in their classroom assessments more. This result is in line with other studies (Arslan and Üçok Atasoy, 2020; Farhady and Tavassoli, 2021; Imran, 2011).

At the surface level, it may be concluded that ELF instructors' overreliance on multiple-choice items may prevent students from developing high-order thinking skills and thorough comprehension (Haladyna et al., 2002) but the interview results might show that the instructors were aware of this problem. According to social constructivist theory, instead than forcing learners to understand texts in accordance with a teacher's perspective, teachers need to encourage them to make their own meaning from the texts (Yang and Wilson, 2006). All of the instructors mentioned that they used open-ended questions and pre-reading sessions to activate the schemata and post-reading sessions or presentations to allow student interaction and collaboration so that instructors would have more chance to monitor students' learning not only on a surface level but also on an intensive level and students would have much more time to practice and learn from each other.

In addition to these, the instructors also indicated that they use the student-conducted method of summarizing which is inextricably tied to extensive reading comprehension, since the student must first make sense of the material (Kintsch and Van Dijk 1978). This result contradicts the findings of Ketabi's (2015) study which found out teachers use summarizing less than other methods. On the other hand, the findings of the same study support that teachers do not use student journals and

portfolios due to the lack of knowledge on how to use them and their workload.

Another note-worthy finding was about the read-aloud method. The read-aloud method was stated to be the least favoured method among the instructors both according to the questionnaire and interview results. The instructors in interviews stated that they don't prefer reading aloud because it is not a useful tool to assess reading competency and it is not authentic. The result of this study contradicts Swai's (2019) findings in which read-aloud was among the main student-conducted methods since teachers believed that pronunciation is a crucial component of reading competency and reading aloud can assist students in improving it. In this study, the reason for not focusing on pronunciation may be because the instructors consider that they use plenty of opportunities to help students improve their pronunciation skills in integrated skill, listening and speaking lessons. It may also be linked to the fact that the instructors were aware that reading aloud more serves to decoding rather than meaning-making.

Writing Assessment: Regarding writing assessment, the dominant methods were instructor-made assessments and standardized tests. The most preferred instructor-made assessment methods were indicated as short essays, editing a sentence and a paragraph and long essays according to the questionnaire results, whereas multiple-choice and matching items were the least favoured ones. The questionnaire results were also in line with the interview results in which instructors stated that they use short paragraphs, and short or long essays according to the proficiency level of students. The results pointed out that the instructors mostly adopt direct, responsive and extensive writing methods. The instructors' preference for the responsive tasks was in line with the findings of Ayas' (2014) study which found that teacher-made tests were the most favored assessment methods. The preference for essays or paragraphs can be associated with the nature of writing skills. As writing is a productive skill, a direct assessment which is the assessment of students' skills based on their products or performances (Fraidan and Mulhim, 2012), is a better option as direct assessment corporates all aspects of writing such as content, accuracy, sentence order, and organization, and assesses students' communicative competence (Coombe et al., 2007). Moreover, essays also require students to develop critical thinking skills. This is reflected in the interviews by the instructors as a disadvantage of essay writing as they need to spend a lot of time on pre-writing

sessions as students are not able to find ideas.

The standardized tests, as in reading assessment, were the second favoured assessment method according to the interview results. Although the instructors didn't mention standardized tests as their popular method and they underlined that they are not grading students' work, they seemed to be aware of preparing their students for the exams as they are still the dominant methods in their institution. Because of this reason, the washback effect is reflected in their comments. For instance, one instructor mentioned the students need to practice until the "real" one. Some teachers mentioned that they showed the criteria to their students because they are responsible for the same criteria in the exam. However, when the teachers concentrate all of their lesson plans on the subjects and skills that will be assessed on a given test, they might overlook other crucial facets of learning and narrow the curriculum (Cooley, 1991).

As for the student-conducted methods, the findings of interview and questionnaire results showed self and peer assessment were adopted less than instructor-made methods and standardized tests. However, considering that more than half of the instructors stated they "often" or "usually" use them implies that they are aware of its potential implications in promoting self-regulation skills. The results of Khodadady and Khodabakhshzade (2012)'s study on the effect of self-assessment on writing ability and autonomy, also showed that self-assessment improved both students' self-regulation and writing skills. The interview results also revealed although the instructors use peer assessment to enable students to learn from each other by editing their writings, they also find it challenging which is reflected in their comments about the difficulty of carrying out peer assessment, especially with lower-level students. In contrast to self and peer assessment, student journals were the least preferred method among the instructors. The reason may be linked to the instructors' practices and knowledge. As the instructors do not reflect on their teaching themselves, they may regard student journals which let students reflect on the writing process as time-consuming. Also similar to their beliefs on the read-aloud method, they may not prefer student journals as they are not directly related to outcome or language production. Al-Nouh et al. (2014) investigated EFL Primary School Teachers' Attitudes, Knowledge and Skills in Alternative Assessment. The finding of this study supports this claim by revealing, teachers' preference of

traditional methods over alternative ones as they take a lot of time and overlook students' writing skills directly.

Listening and Speaking Assessment: The information gathered from the surveys and interviews provides an in-depth understanding of the methods the instructors employ for assessing students' speaking and listening skills. Unlike the results of reading and writing, for listening and speaking, the instructors preferred student-conducted methods more for assessing listening and speaking. The questionnaire data showed that the instructors prefer oral interviews and oral discussions more than other methods. This finding was parallel to Saefurrohman's (2018) result according to which respondents mostly assessed their speaking and listening skills through oral interviews/dialogues, with a mean score of 3.62, 62.50% of respondents consistently utilized it. Parallel with this result, Luoma (2004) pointed out that oral interviews have been the most frequently utilized method since the beginning of the communicative language teaching era. This may be due to the fact that interviews provide a more authentic environment to assess students' listening and speaking competence. On the other hand, during interviews, oral dialogues were not indicated to be a popular method among instructors. One instructor highlighted students tended to memorize when they were practicing dialogues. The instructors' inclination may result from a desire to assess real-world conversations when they talk spontaneously and negotiate the meaning. The interview findings hinted that the instructors value assessment methods which include student engagement more. Some instructors also stated that they try to include collaboration in listening activities. Assessing listening and speaking skills by combining techniques like jigsaw listening and peer assessment with summarizing and retelling stories may promote critical thinking, active participation, and group learning which are all crucial for language learning. Vygotsky's sociocultural theory supports that claim by highlighting learning is most successful when individuals communicate and share their knowledge with each other (Nunan, 1999). The research conducted by Idham et al. (2024) was also in line with this by showing interviews improve EFL students' speaking abilities, leading to more active participation in learning and increase confidence in English communication. Also, the findings of Tavi's (2010) study, which is about the effect of integrating listening and speaking skills on the communicative competence of students, showed that the group practicing integrated skills performed better than

those who practiced listening and speaking separately.

Another method indicated to be used frequently by the instructors according to interview results was summarizing. The instructors stated that they combine retelling and summarizing methods. One of the instructors stated that s/he uses this method to give students a purpose to listen. Unlike multiple-choice items which primarily involved lower-level cognitive processing, summarizing and retelling require listeners to use higher-level cognitive processing like meaning and discourse construction as well as metacognitive strategies. While cognitive processing occurs automatically, listeners can learn metacognitive strategies to control and monitor their listening processes. That way, teachers can monitor students' learning process of skills, rather than focusing on test-taking strategies (Rukthong, 2021).

Following student-conducted methods, the most popular method was the instructor-made method of multiple-choice test items following a listening passage according to the interview data. The findings of the study conducted by Saefurrohman (2018) to investigate EFL teachers' assessment methods in oral communications also found that multiple-choice exam questions after a listening passage was the most preferred teacher-made method among Indonesian English instructors. However, the multiple-choice method does not let students synthesize and demonstrate their knowledge; instead, it places more emphasis on selecting the right response from a list of possibilities through recalling. Thus, it neither properly fosters the development or assessment of skills nor does it encourage the use of such skills in various contexts (Paxton, 2000). The reason was revealed in the comments of interviewers as the effect of books and testing offices and exams. That curriculum office is accountable for determining the learning outcomes might cause instructors to rigorously adhere to the curriculum, which frequently limits methods for assessment. Another reason may be as the testing offices determine the type of tasks for exams, the instructors may need to create extra tasks themselves. Some of them may find it difficult because of the workload and rely heavily on multiple-choice types. This claim was also supported by another instructor in the interview which showed that some instructors include open-ended questions to promote deeper learning and critical thinking. The last reason may be the negative washback effect of exams on students' motivation. They may be students' being uninterested in listening and speaking activities when the instructors do not include the standardized testing

methods. For instance, Instructor 9 stated that since speaking and listening abilities are not assessed, students were less motivated to participate in lessons.

In addition to multiple-choice items, most of the instructors in the interview highlighted the importance of the note-taking method (especially the Cornell method) and counted it as their most favoured method. Based on the interview data, factors that contributed to teachers' high use of the Cornell note-taking method were, preparing students for their departments, the positive effect of note-taking on students' active listening and comprehension and washback. First of all, note-taking is quite useful since it simulates actual listening situations where taking notes is necessary. As Instructor 5 mentions, after preparation classes, students may need to listen to the lectures in English in their departments and take notes. The instructors may be trying to prepare students for their departments. Secondly, to take notes, students need to listen carefully. This may enhance their comprehension. In Cheng's (2004) study similarly, taking notes and students' summaries were the most popular two methods. The data showed that the percentage of teachers using note-taking in Canada (70,3 %) and Beijing (70,5%) was high. A study conducted by Hayati and Jalilifar (2009) showed a strong correlation between note-taking techniques and listening comprehension proficiency. Furthermore, students who used their own note-taking technique performed at a lower language performance level than those who used the Cornell approach.

Alongside the assessment methods, the interview data also shed light on the importance of the level of students in the selection and difficulty of tasks. Less proficient students' tasks were indicated to be simpler and directly connected to the material covered in class or their daily lives. Lower-level students are also provided with thematic questions and speaking strategies to get the necessary input until they have autonomy. The instructors' inclination may be related to Krashen's Input Hypothesis which highlights the importance of receiving comprehensible input to acquire a language (Krashen, 2010).

D. What are EFL Instructors' Beliefs about Classroom Assessment? To What Extent are the Beliefs of Instructors in Line with Their Practices?

Seven questionnaire items regarding the beliefs of the instructors are based on the principles of AfL. Understanding the connection between teaching and

learning—more especially, how to use assessment to direct instruction—is the first step in achieving AfL (Chen, 2020). Since "social constructivism" is the foundation of the learning theory approach that underpins AfL, learners actively participate in the process, and the assessment has a formative role (Stobart, 2008). This approach seeks to maintain a balance between the individual and cultural components of learning. Criterion-reference assessment and descriptive feedback are therefore the key elements of this process. Rather than comparing students' scores with each other, teachers negotiate learning outcomes and success criteria and provide samples of assessment to see if the criteria are met or if the learning outcomes are fulfilled. To achieve this, learners need to understand the standards that are aimed for (Stiggins and Chappuis, 2005). While instruction is taking place, teachers constantly search for evidence of how student learning is progressing toward a goal. Descriptive feedback and criterion-reference assessment are thus the essential components of this procedure. Teachers negotiate learning outcomes and success criteria and present assessment samples to determine if the criteria are met or if the learning outcomes are achieved, as opposed to comparing students' results with one another. To accomplish this, students must comprehend the criteria being pursued (Stiggins and Chappuis, 2005).

According to the questionnaire results, given that the mean scores for assessments in the classroom were greater than 4.00 and the majority of instructors strongly agreed or agreed with every statement except the last one, the findings of instructors' perspectives of assessments showed that almost all instructors had a strong tendency to employ AfL.

When each item of the questionnaire was examined, the data showed that the item with the highest mean score was the statement *"Students should be given feedback which enables them to know how to improve their work and take their learning forward."* The majority of educators thought that students' progress may be positively impacted by feedback. First off, 97% of teachers agreed—strongly—or agreed—that it is essential to provide feedback to students so they can learn how to improve their work and progress in their studies.

The interview data revealed that the beliefs or instructors were reflected in their practices in terms of giving feedback. First of all, in the interviews, the instructors indicated that they use various forms of feedback. While written and oral

feedback were stated to be the most preferred ones, some instructors also mentioned they use conferencing with students. This result is also similar to Fawzi and Alldabous' (2019) study which showed that teachers use both oral and written feedback more than other forms. Furthermore, the instructors highlighted that it is vital to give descriptive feedback rather than showing them their grades to enable students to see their weaknesses and to have a chance to work on them. For instance, during speaking activities, one of them mentioned that they monitor students during speaking activities and provide real-time feedback. Also, another instructor stated that in writing, they use process-feedback which involves more than one draft, by writing symbols and comments to allow students to reflect on their work and improve it. Parallel with this claim, according to Black and William (1998) effective feedback focuses on specific features of students' work and offers suggestions for improvement, not on the comparison of students with each other. In addition to timely feedback, the interview data also revealed that the instructors tend to balance positive and negative feedback to increase students' motivation. However, the interview data did not show how teachers guide students to improve their work or how they give opportunities to them to close the gap.

The second item with the highest mean score was about identifying the purpose of assessment when designing tasks. In the questionnaire results about the purpose of assessment, 74% of the instructors strongly agreed on that item. In the interviews, they also emphasized that they try to be sure that their assessment tasks match with learning outcomes or syllabi. For instance, Instructor 3 mentioned that s/he decides on the type of feedback according to the purpose of the activity, focusing on either fluency or accuracy. Also, Instructor 7 stated that they ensure that the students achieve the learning outcomes provided by the curriculum office.

Furthermore, the interview data for all methods shows that although the majority of instructors (n=94) either strongly agreed or agreed with the statement that *"Students should be given opportunities to reflect on their learning through peer feedback and self-assessment."* and most of them (n=89) agreed or agreed that it has an effect on their achievement and motivation when the interview results regarding the methods were taken into consideration, it can be seen that they practice it at a moderate level for all skills. The interview results showed that while seven instructors indicated that they use self and peer assessment in the class, three of them

pointed out the difficulty of utilizing it due to students' proficiency levels and time constraints. Similarly, Balboa Álvarez (2020) conducted a research on teachers' beliefs on peer and self-assessment in the EFL upper-secondary school classroom in Sweden. The findings showed that although the questionnaire data supported teacher's preference for these methods, the interviews did not show consistency between their beliefs and practices.

The last item about the beliefs of the instructors on AfL was about whether the instructors use assessment results to modify teaching and learning. The interview data showed that many instructors (n=93) strongly agreed or agreed that in order to best support students' learning, instruction should be modified based on assessment results. Also, in the questionnaire about the purposes of assessment, the item related to diagnosing their own strengths and weaknesses in teaching was the second most rated teacher-related purpose. The interview data also supported that the instructors modify their instruction and provide remedial teaching to tailor to students' needs. They also stated that they share the assessment results with other instructors to modify the curriculum in case of need. On the other hand, two instructors mentioned the difficulty of modifying their teaching due to tight schedules and structural constraints. Starting from this result, it can be said that the instructors are aware of the role of assessment to improve teaching.

The last item with a strong consensus (n=88) among the instructors was about providing students with assessment criteria in advance. In the interviews, the instructors stated that they not only share the assessment criteria before the tasks but also clarify assessment criteria by providing examples for students to understand what is expected from them. When the instructors use criterion such as learning objectives or syllabus-related objectives, students can be more motivated as they will be able to work at their own pace, rather than competing with other students. Instructors can also see students' progress in a goal-based environment and identify what they achieved or if they have any learning difficulties (Stiggins and Chappuis, 2005). This shows that by providing assessment criteria, the instructor's beliefs are parallel with their practices. This result was similar to the results of Zulaiha et al.'s (2020) study. Although the questionnaire result showed that not many teachers believed that they need to inform students in a detailed way about the assessment criteria, in the interviews, the majority of teachers stated that in order for students to

prepare and provide their best effort, they should be informed about the criteria they would be assessed on. The findings contradicted the results of Widiastuti's (2020) study which revealed that although the interview results showed that the teachers indicated that they modified their instruction, during class observations, it was observed that teachers were not able to provide remedial teaching to students due to time allocation constraints and being busy with administrative duties or the remedial teaching was below the standards. However, as this study was based solely on the data from interviews, the results may not reflect the real situation.

Lastly, even though the interview data showed that the instructors had a high belief in the effect of assessment to improve learning, it was seen that except for listening and speaking, they mostly relied on instructor-made methods and standardized tests.

E. Conclusion

This current research aimed to investigate EFL Instructors' purposes, methods, and beliefs of classroom assessment. Through the use of semi-structured interviews and a 5 Likert scale questionnaire, the study employed a mixed-methods strategy to gather both qualitative and quantitative data.

The first research question looked at the purposes of EFL instructors in classroom assessment. According to the results, EFL instructors use classroom assessment mainly for student-centred, formative purposes: (1) to learn more about students' growth, (2) to give feedback to students during the course, (3) to diagnose the strengths and weaknesses of their students and (4) to motivate students. The interview data also illustrated that the instructors also use the summative purpose of (5) grouping students. Although in the questionnaire the instructional purpose of diagnosing strengths and weaknesses in teaching is selected by many instructors, the interview data revealed a discrepancy between the questionnaire and interview results. Very few respondents said in the interviews that they use assessment to identify their own strengths and weaknesses which could be linked to an overemphasis on student learning and lack of sufficient training on how to use assessment to reflect on their own teaching.

The second research question focused on the methods used by EFL

instructors to assess the four language skills: reading, writing, listening, and speaking. Although the literature suggests using both summative and formative assessment, to assess reading skills, according to the questionnaire and interview results, overall, instructors tend to employ standardized examinations and instructor-made methods for reading and writing. The most common methods used for assessing reading were multiple-choice, true/false, and matching items and student summaries while the least favoured method was reading aloud since it is not directly related to reading comprehension and it is not authentic. Standardized exams like quizzes, midterms and finals were dominant forms of reading assessment. The interview results also revealed that while instructors are aware of the constraints of these methods in promoting high-order skills, these methods are utilized due to practicality, reliability, and the school policy therefore, instructors have no other choice most of the time. Furthermore, as the instructors are aware that their students will be tested on those items, they depend more on them. Despite these limitations, the interview data uncovered that instructors incorporate some other methods into the classroom assessment such as asking open-ended questions in pre-while and post-reading stages, summaries, or presentations to foster creativity, and collaboration and monitor students' learning. To assess writing, instructors have a similar inclination toward summative, instructor-made methods. The most frequently used methods for assessing writing were short essays, editing a sentence and a paragraph and long essays according to the interview results while multiple-choice, matching items and student journals were the least favoured ones, demonstrating a preference for extensive, direct, and responsive writing methods. While the instructors use instructor-made methods, from the interview data it is obvious that they use formative approaches to the assessment of writing. They indicated that they share criteria with students, provide samples for strong and weak works to make students learn what quality work looks like and include students in peer-assessment practices. Finally, the data results showed that for assessing listening & speaking, instructors use both student-conducted methods of multiple-choice items, taking notes and summarizing for listening. In the interview, it is revealed when students are not formally assessed in listening skills, they lose their interest in lessons. Also, the dominant effect of the Testing Office and Curriculum forced instructors to rely on multiple-choice items. While the instructors use more instructor-made methods for assessing listening skills, for speaking students conducted methods of oral

presentations, oral interviews and oral discussion were utilized more by the instructors. The need for real-life interaction and authentic communication, rooted in social constructivist theory and communicative language teaching encourages instructors to use more authentic assessment methods to assess listening and speaking. However, from the interview results, it is obvious that the instructors' decisions are affected by institutional policies that impose using reliable and practical methods and their own preferences which are inclined to using authentic assessment methods and preparing students for real-life skills.

Finally, the third research question was to explore instructors' beliefs on the AfL function of classroom assessment and to see to what extent their beliefs were in line with their practices. According to the questionnaire data, the instructors have positive beliefs on the items related to the AfL. The information obtained from the interview shed light on the instructor's views regarding criteria, feedback, peer-self assessment, and the intended use of assessment results in the classroom. Findings of the questionnaire showed that the instructors had strong beliefs on AfL and they reported that they mostly put AfL principles into practice in the areas of providing feedback and using assessment results to modify instruction. On the other hand, their beliefs were not truly aligned with their practices on methods, student involvement through self and peer assessment practices and reflecting on their teaching due to challenges such as time constraints, administrative limitations, tight schedules, negative effects of exams, and students' low proficiency level. Due to the limitations of interview results, and lack of observation results, it is hard to come to a definitive conclusion about the extent to which the instructors' beliefs align with their practices especially in terms of examining the practices of instructors to guide students to close the gap. Though the interview data offer insightful information about the instructors' practices, it lacks the real classroom observations. A more through picture can only be captured by observing the real classroom interactions.

F. Limitations of the Study

This study had three major limitations. The first limitation was related to the small sample size. As 100 instructors working at preparation schools of universities in Istanbul, İzmir, Bursa, Aydın, Mersin, Ankara and Eskişehir participated in the study, the results cannot be generalized to the target population.

The second limitation was related to the interviews. 10 participants were accepted to attend the interviews. Even though the interviews helped clarification of some points for the questionnaires, a larger number of respondents may have disclosed more information and maybe offered greater clarification of the questionnaire answers. Furthermore, in addition to interviews, classroom observations are needed to learn more about the instructors' real practices in the classroom and to see the diversity of methods used in classroom assessment. Finally, the assessment literacy of instructors who use classroom assessment needs further investigation.

G. Suggestions for Further Research

The current research aimed at exploring the alignment of instructors' practices with their beliefs and practices. Considering that the research depending on the interview data created a limitation, further research which use classroom observations, together with instructors' journals could help us gain a better understanding of instructors' classroom assessment practices.

Furthermore, the participants of this study consist of 100 instructors from Istanbul, İzmir, Bursa, Aydın, Mersin, Ankara and Eskişehir. The generalizability of the results can be enhanced by expanding the sample size and making sure that teachers are diversely represented from different areas, school types (public vs. private), and educational levels (primary, secondary, tertiary).

Thirdly, the most commonly stated factors affecting the practices of instructors was related to institutions policies, testing and curriculum. Consequently, it is critical to look into how these institutional components influence and limit EFL instructors' assessment practices. Similarly, the effect of policy makers may be investigated to see the effect of national policies on the classroom assessment practices.

In addition to these, this study only examined instructors' beliefs and practices of classroom assessment. Future studies might also look into how students feel about the methods and approaches their teachers employ for assessment in the classroom, as well as how they perceive them. This could assist shed light on the role of assessment in promoting students' learning especially in terms of showing them

where they are, where they desire to be and how they close the gap.



VI. REFERENCES

BOOKS

- AFFLERBACH, P., and CHO, B. (2015b). "The Classroom Assessment of Reading", In KAMIL, M. L., PEARSON, D. P., MOJE, E. B., AFFLERBACH P. (Eds.) **Handbook of Reading Research**, Routledge.
- AIRASIAN, P. and RUSSELL, M. (2008). **Classroom Assessment: Concepts and Applications**, McGraw-Hill Higher Education.
- ALMARODE, J., FISHER, D., FREY, N., and THUNDER, K. (2021). **The Success Criteria Playbook: A Hands-On Guide to Making Learning Visible and Measurable**, Corwin Press
- BACHMAN, L. F. and PALMER, A. S. (1996). **Language Testing in Practice: Designing and Developing Useful Language Tests**, Oxford University Press.
- BAILEY, K. (1988). **Learning about Language Assessment: Dilemmas, Decisions, and Directions**, Boston, Heinle and Heinle.
- BLACK, P. J. and WILLIAM, D. (1998). **Inside the Black Box Raising Standards Through Classroom Assessment**, Granada Learning.
- BLOXHAM, S., and BOYD, P. (2007). **Developing Effective Assessment in Higher Education: A Practical Guide**. New York, NY, McGraw-Hill International.
- BREEN, M. P., BARRATT-PUGH, C., DEREWIANKA, B., HOUSE, H., HUDSON, C., LUMLEY, T. and ROHL, M. (1997). **Profiling ESL Children: How Teachers Interpret and Use National and State Assessment Frameworks**, Canberra: Department of Employment, Education, Training and Youth Affairs.

- BROOKHART, S. M. (2017). **How to Give Effective Feedback to Your Students**, ASCD.
- BROWN, A. L. (1985) “Metacognition: The Development of Selective Attention Strategies for Learning from Texts”, In SINGER, H and RUDELL, R. B. (Eds.) **Theoretical Models and Processes of Reading**, Newark: International Reading Association.
- BROWN, H. D. (2004). **Language assessment: Principles and Classroom Practices**, New York: Pearson/Longman.
- BROWN, G., and YULE, G. (1983). **Teaching the Spoken Language**. Cambridge: Cambridge University Press.
- CAMP, R. (1993). “The Place of Portfolios in Our Changing Views of Writing Assessment”, In BENNET, R. E and WARDS, W. C. (Eds.), **Construction Versus Choice in Cognitive Measurement**, Routledge.
- CARLESS, D., and WINSTONE, N. (2019). **Designing Effective Feedback Processes in Higher Education: A Learning-Focused Approach**, London, UK, Routledge.
- CARR, N. T. (2011). **Designing and Analyzing Language Tests**. Oxford, Oxford University Press.
- CARROL, B. J. (1980). **Testing Communicative Performance**, London: Pergamon.
- CHENG, L., and FOX, J. (2017). **Assessment in the Language Classroom: Teachers Supporting Student Learning**, Red Globe Press.
- CLARKE, C. (2004). **Department for Education Skills: Five-Year Strategy for Children and Learners**, Norwich, HMSO
- COOMBE, C. (2018). **An A to Z of Second Language Assessment: How Language Teachers Understand Assessment Concepts**, London, UK, British Council.
- COOMBE, C., DAVIDSON, P., O’SULLIVAN, B. and STOYNOFF, S. (2012). **The Cambridge Guide to Second Language Assessment**, New York, Cambridge University Press.

- COOMBE C., FOLSE K., and HUBLEY N. (2007). **A Practical Guide to Assessing English Language Learners**, Michigan, The University of Michigan Press.
- CROCKER, L. M., ALGINA, J. (2008). **Introduction to Classical and Modern Test Theory**, Cengage Learning.
- DOLIN, J., BLACK, P., HARLEN, W., and TIBERGHIE, A. (2018). “Exploring Relations Between Formative and Summative Assessment”, In J. Dolin and R. Evans (Eds.), **Transforming Assessment**, Switzerland, Springer International Publishing.
- EARL, L. (2003). **Assessment as Learning: Using Classroom Assessment to Maximise Student Learning**, Thousand Oaks, Corwin Press.
- FRANKLIN, J. (2002). “Assessing Assessment: Are Alternative Methods Making the Grade” In **Curriculum Update**, ASCD.
- FRAENKEL, J. R., WALLEN, N. E. and HYUN, H. H. (2012). **How to Design and Evaluate Research in Education**. New York, McGraw-Hill Education.
- FULCHER, G. and DAVIDSON, F. (2007). **Language Testing and Assessment**, London, Routledge.
- GARCÍA JIMÉNEZ, E., GALLEGU NOCHE, B., and Y GÓMEZ RUIZ, M. A. (2015). “Feedback and Self-Regulated Learning: How Feedback Can Contribute to Increase Students’ Autonomy as Learners”, In PERIS-ORTIZ, M. and MERIGÓ LINDAHL, J. M. (Eds.), **Sustainable Learning in Higher Education, Innovation, Technology, and Knowledge Management**, London, Springer.
- GUILFORD, J. P. (1954). **Psychometric methods**, New York: McGraw-Hill.
- HAMP-LYONS L. and CON W. (2000). **Assessing the Portfolio Principles for Practice, Theory and Research**, Hampton Press.
- HARRIS, D.P. (1969). **Testing English as a Second Language**, New York, McGraw-Hill.
- HARMER, J. (2017). **The Practice of English Language Teaching**. London,

Longman Group.

HATİPOĞLU, Ç. (2021). "Testing and Assessment of Speaking Skill, Test Task Types, and Sample Test Items", In ÇELİK, S., ÇELİK, H. and Coombe C. (Eds.), **Language Assessment and Test Preparation in English as a Foreign Language (EFL) Education**, Vizetek.

HEATON, J. B. (1988). **Writing English Language Tests**, Longman.

HIDRI, S. (2018). "Discrete Point and Integrative Testing" In LIONTAS J.I. (Ed.), **The TESOL Encyclopedia of English Language Teaching**. John Wiley and Sons.

HUBA, M. and FREED, J. E. (2000). **Learner-centered Assessment on College Campus: Shifting the Focus from Teaching to Learning**, Pearson.

HUBLEY, N. J, "Reading Assessment", In. LIONTAS, J. I., INTERNATIONAL ASSOCIATION, and DELLICARPINI M. (Eds.), **The TESOL Encyclopaedia of English Language Teaching**. DOI: <https://doi.org/10.1002/9781118784235.eelt0336>.

HUGHES, A. (1989). **Testing for Language Instructors**, Cambridge, Cambridge university press.

KRASHEN, S (2010). "The Goodman/Smith Hypothesis, the Input Hypothesis, the Comprehension Hypothesis, and the (Even Stronger) Case for Free Voluntary Reading" In ANDERS, P. (Edt.), **Defying Convention, Inventing the Future in Literary Research and Practice**, New York: Routledge.

KUNNAN, A. and JANG, E. E. (2009). "Diagnostic Feedback in Language Assessment", In LONG M. H. and DOUGHTY C. J. (Eds.), **The Handbook of Language Teaching**, Wiley-Blackwell.

LADO, R. (1961). **Language Testing: The Construction and Use of Foreign Language Tests**, London, Longman.

LAM, R. (2018). "Portfolio Assessment for the Teaching and Learning of Writing", In JIANG H. and HILL M.F. (Eds.), **Teacher Learning from**

- Classroom Assessment: Perspectives from Asia Pacific**, New York, Springer.
- LAMMERS W. J. and BADIA P. (2005). **Fundamentals of Behavioral Research**, Thomson, Wadsworth.
- LAMBERT, D., and LINES, D. (2000). **Understanding Assessment**. London, Routledge Falmer.
- LINDSTROM, C., and DROLET, B. M. (2017). **What's Missing: Best Practices for Teaching Students with Disabilities**, Rowman and Littlefield.
- LOUMA, S. (2004). **Assessing Speaking**, Cambridge University Press.
- MADSEN, H. S. (1983). **Techniques in Testing**, Oxford University Press.
- MCMILLAN, J. H. (2013). **Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction**, Boston, Pearson.
- MCNAMARA, T. (2000). **Language Testing**. Oxford Press.
- MILLER, M.D., LINN, R.L. and GROUNDLUND, N.E. (2008). **Measurement and Assessment in Teaching**, Pearson International Edition.
- MILLAR, R., and HAMES, V. (2006). "Using research to clarify learning goals and measure outcomes", In **Improving Subject Teaching**, Routledge.
- NAKATSUHARA F., KHABBAZBASHI N., and INOUE C. (2022). "Assessing Speaking", in FULCHER G., HARDING L. (Eds), **Routledge Handbook of Language Testing**, Routledge.
- NUNAN, D. (1999). **Second Language Teaching and Learning**, Boston, MA, Heinle and Heinle.
- O'MALLEY, J., and PIERCE, L. (1996). **Authentic Assessment for English Language Learners: Practical Approaches for Teachers**. New York, Pearson.
- ORNSTEIN, Allan C. and HUNKINS, F. P. (2017). **Curriculum: Foundations, Principles, and Issues**, Pearson Higher Education.

- PAI, J. (2018). "Observations and Conversations as Assessment in Secondary Mathematics", In THOMPSON, D., BURTON, M., CUSI, A. and WRIGHT, D., **Classroom Assessment in Mathematics: Perspectives from Around the Globe**, Springer.
- PANADERO, E., LIPNEVICH, A., and BROADBENT, J. (2019). "Turning Self-Assessment into Self-Feedback", In HENDERSON, M., AJJAWI, R., BOUD, D., and MOLLAY, E. (Eds.), **The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners**, Palgrave, Macmillan.
- PARIS, S. C., and STAHL, S. A. (2005). **Children's Reading Comprehension and Assessment**. London: Lawrence Erlbaum Associates Publishers.
- PAWLAK, M., and WANIEK-KLIMCZAK, E. (2015). **Issues in Teaching, Learning and Testing Speaking in A Second Language**, Springer.
- POPHAM, J.W. (2009). "Formative Assessment: Monitoring Instruction and Learning: A Chapter from Instruction That Measures Up". In POPHAM, J. W. (Ed.) **Instruction That Measures Up: Successful Teaching in The Age of Accountability**, ASCD.
- POPHAM, J.W. (2014). **Classroom Assessment: What Teachers Need to Know**, Boston: Pearson.
- PRITCHARD, A., and WOOLLARD, J. (2010). **Psychology for the Classroom Constructivism and Social Learning**, Routledge.
- PUNCH, K. (2005). **Introduction to Social Research: Quantitative and Qualitative Approaches** (2nd Edition), London, Sage.
- RICHARD, J. C. and SCHMIDT, R. (2002). **Longman Dictionary of Language Teaching and Applied Linguistics**, London: Pearson Education.
- ROST, M. (2013). **Teaching and Researching: Listening**, Routledge.
- RUBIN, J. and THOMPSON, I. (1994). **How to Be A More Successful Language Learner**, Boston, Heinle and Heinle.
- SHARON L. NICHOLS. (2021). **Teaching on Assessment**, Information Age Publishing.

- SHERMIS, M. D. and DIVESTA, F. J. (2011), **Classroom Assessment in Action**, Rowman and Littlefield Publishers.
- STIGGINS, R. J. (2010). “Conquering the Formative Assessment Frontier”, In MCMILLAN (Ed.), **Formative Classroom Assessment**, New York, Teachers College, Columbia University.
- STOBART, G. (2008). **Testing Times: The Uses and Abuses of Assessment**, London, Routledge.
- TURNER, C. E. (2012). “Classroom Assessment”, In FULCHER, G. and DAVIDSON, F. (Eds.), **The Routledge Handbook of Language Testing**, New York, NY/: Routledge.
- VALENCIA, S.W, HIEBERT, E. H., and AFFLERBACH P.P., (1994). **Authentic Reading Assessment: Practices and Possibilities**, International Reading Association.
- WEIGLE, S. C. (2002). **Assessing Writing**. Cambridge, Cambridge University Press.
- WEIR, C. J. (1990), **Communicative Language Testing**, New York, Prentice Hall.
- WIGGINS, G. P. (1993). **Assessing Student Performance: Exploring the Purpose and Limits of Testing**, Jossey-Bass/Wiley.

ARTICLES

- ABEYWICKRAMA, P. (2012). “Rethinking Traditional Assessment Concepts in Classroom-Based Assessment”, **CATESOL Journal**, Volume 23, Issue 1, Pages 205-213.
- ACAR ERDOL, T. and YILDIZLI, H. (2018). “Classroom Assessment Practices of Teachers in Turkey”, **International Journal of Instruction**, Volume 11, Issue 3, Pages 587-602. DOI: <https://doi.org/10.12973/iji.2018.11340a>.
- ACAROL, K. (2024). “The Effect of Web-Based Peer Feedback on Students’ Writing Achievement”, **Journal of Educational Technology & Online Learning**, Volume 7, Issue 1, Pages 52-70.
DOI: <https://doi.org/10.31681/jetol.1314382>.

- ADEOYE-OLATUNDE, A. and OLENIK, N. L (2021). Research and scholarly methods: Semi-structured interviews. *Journal of the American College of Clinical Pharmacy*, Volume 4, Issue 10, Pages 1358-1367. DOI: <https://doi.org/10.1002/jac5.1441>.
- ADNAN, A., SUWANDI, S., NURKAMTO, J., and SETIAWAN, B. (2019). "Teacher Competence in Authentic and Integrative Assessment in Indonesian Language Learning". **International Journal of Instruction**, Volume 12, Issue 1, Pages 701–716. DOI: <https://doi.org/10.29333/iji.2019.12145a>.
- AHMAD, S., SULTANA, N., and JAMIL, S. (2020). "A Paradigm Shift from Traditional to Alternative Assessment Techniques", **Journal of Applied Linguistics and Language Research**, Volume 7, Issue 2, Pages 19-33.
- AHMADI, A., and SADEGHI, E. (2016). "Assessing English Language Learners' Oral Performance: A Comparison of Monologue, Interview, and Group Oral Test", **Language Assessment Quarterly**, Volume 13, Issue 4, Pages 341-358. DOI: <https://doi.org/10.1080/15434303.2016.1236797>.
- AHMED, M. K. (2018). "Pedagogy in Speaking: Challenges Addressed by Teacher-Student in the ESL Context", **International Journal of Applied Linguistics and English Literature**, Volume 7, Issue 3, Pages 97-107. DOI: <https://doi.org/10.7575/aiac.ijalel.v.7n.3p.97>.
- ALANAZI, A. (2016). "A Critical Review of Constructivist Theory and Constructivism", **American Research Journal of Humanities and Social Sciences**, Volume 2, Pages 1-8. DOI: <https://doi.org/10.21694/2378-7031.16018>.
- ALDERSON, J. C. and WALL, D. (1993). "Does Washback Exist?", **Applied Linguistics**, Volume 14, Issue 2, Pages 115-129. DOI: <https://doi.org/10.1093/applin/14.2.115>.
- AL-NOUH, N.A., TAQI, H. A., and ABDUL-KAREEM, M. M. (2014). "EFL Primary School Teachers' Attitudes, Knowledge and Skills in Alternative Assessment". **International Education Studies**, Volume 7, Issue 5, Pages 68-84. DOI: <https://doi.org/10.5539/ies.v7n5p68>.

- AMINEH, R. J., and ASL, H. D. (2015). "Review of Constructivism and Social Constructivism", **Journal of Social Sciences, Literature, and Languages**, Volume 1, Issue 1, Pages 9-16.
- ANDHIKA, T. R. W. (2011). "Assessment in ELT: from Theory to Practice"
Jurnal Ilmiah Bina Bahasa, Volume 4, Issue 2, Pages 91-100.
- AOKI, S. (2014). "Potential of Voice Recording Tools in Language Instruction", **Studies in Applied Linguistics & TESOL**, Volume 14, Issue 2, Pages 128-141. DOI: <https://doi.org/10.7916/salt.v14i2.1309>.
- APRIDAYANI, A., HAN, W., and SAKKANAYOK, K. (2024). "Enhancing English Writing Competence in Higher Education: A Comparative Study of Teacher-Only Assessment Versus Teacher and Student Self-Assessment Approaches". **Asian-Pacific Journal of Second and Foreign Language Education**, Volume 9, Issue 37, Pages 1-20.
DOI: <https://doi.org/10.1186/s40862-024-00263-3>.
- ARSLAN, R. Ş., and ÜÇÖK-ATASOY, M. (2020). An Investigation into EFL Teachers' Assessment of Young Learners of English: Does Practice Match the Policy?", **International Online Journal of Education and Teaching (IOJET)**, Volume 7, Issue 2, Pages 468-484.
- ASAMOAH, D. SHAHRILL, M. and ABDUL LATIF, S. N. (2003). "Towards Developing Classroom Assessment Literacy: Exploring Teachers' Approaches to Assessment Across Cultures", **Cogent Education**, Volume 10, Issue 2, Pages 1-26.
DOI: <https://doi.org/10.1080/2331186X.2023.2280301>.
- ASLANOĞLU, A. E. (2022). "Examining the Effect of Peer and Self -Assessment Practices on Writing Skills", **International Journal of Assessment Tools in Education**, Volume 9, Special Issue, Pages 179-196.
DOI: <https://doi.org/10.21449/ijate.1127815>.
- ATI, A., and PARMAWATI, A. (2022). "The Use of Oral Presentation in Teaching English to Improve Students Speaking Skill", **PROJECT (Professional**

- Journal of English Education**), Volume 5, Issue 2, Pages 300-305.
DOI: <https://doi.org/10.22460/project.v5i2.p300-305>.
- AYUANITA, K. (2013). "Assessing Listening in The Language Classroom",
OKARA: **Jurnal Bahasa Dan Sastra**, Volume 7, Issue 1, Pages 115-122. DOI: <https://doi.org/10.19105/ojbs.v7i1.439>.
- BADEMCI, V. (2006). "To Understand Reliability Properly and to Overthrow Some Clichés: Contrary to The Known Facts, Cronbach's Coefficient Alpha Can Be Negative and Smaller Than "-1". **İnönü University Journal of the Faculty of Education**, Volume 7, Issue 12, Pages 3- 26.
- BİRGİN, O. and BAKI, A. (2007). "The Use of Portfolio to Assess Student's Performance", **Journal of Turkish science education**, Volume 4, Issue 2, Pages 75-90.
- BIRJANDI, P., and TAMJID, N. H. (2010). "The Role of Self-Assessment in Promoting Iranian EFL Learners' Motivation", **English Language Teaching**, Volume 3, Issue 3, Pages 211-220.
DOI: <https://doi.org/10.5539/elt.v3n3p211>.
- BLACK, P., and WILIAM, D. (2009). "Developing the Theory of Formative Assessment", **Educational Assessment, Evaluation and Accountability**, Volume 21, Issue 1, Pages 5–31. DOI: <https://doi.org/10.1007/s11092-008-9068-5>.
- BOSTON, C. (2002). "The Concept of Formative Assessment. Practical Assessment", **Research & Evaluation**, Volume 8, Issue 9, Pages 1-4.
DOI: <https://doi.org/10.7275/kmcq-dj31>.
- BOUBRIS, A. A., and HADDAM, F. (2020). "Reading Assessment: A Case Study of Teachers' Beliefs and Classroom Evaluative Practices, **Arab World English Journal**, Volume 11, Issue 4, Pages 236–253.
DOI: <https://doi.org/10.24093/awej/vol11no4.16>
- BOUD, D. (2000). "Sustainable Assessment: Rethinking Assessment for the Learning Society", **Studies in Continuing Education**, Volume 22, Issue 2, Pages 151-167. DOI: <https://doi.org/10.1080/713695728>

- BROOKHART, S. M. (2008). "Feedback That Fits", **Educational Leadership**, Volume 65, Issue 4, Pages 54-59.
- BROOKHARD, S. M. and DURKIN, D. T. (2003). "Classroom Assessment, Student Motivation, and Achievement in High School Social Studies Classes", **Applied Measurement in Education**, Volume 16, Issue 1, Pages 27-54.
DOI: https://doi.org/10.1207/S15324818AME1601_2.
- BROOKS, L. A. (2009). "Interacting In Pairs In A Test of Oral Proficiency: Co-Constructing A Better Performance", **Language Testing**, Volume 26, Issue 3, Pages 341– 66.
DOI: <https://doi.org/10.1177/0265532209104666>.
- BUTLER, D. and WINNE, P. (1995). "Feedback and Self-Regulated Learning: A Theoretical Synthesis", **Review of Educational Research**, Volume 65, Issue 3, Pages 245-281.
DOI: <https://doi.org/10.3102/00346543065003245>.
- CEYHAN, S., and YILDIZ, M. (2021). "The Effect of Interactive Reading Aloud on Student Reading Comprehension, Reading Motivation and Reading Fluency", **International Electronic Journal of Elementary Education**, Volume 13, Issue 4, Pages 421–431.
DOI: <https://doi.org/10.26822/iejee.2021.201>.
- CHEN, C.W. (2016). "A Survey on EFL Teachers' Assessment Methods in Entry-Level Writing Courses in Technological Universities in Taiwan", **Journal of Pan-Pacific Association of Applied Linguistics**, Volume 20, Issue 1, Pages 21-36.
- CHEN, P. P. and BONNER, S. M. (2020). "A Framework for Classroom Assessment, Learning, And Self-Regulation", **Assessment in Education: Principles, Policy and Practice**, Volume 27, Issue 4, Pages 373-393.
DOI: <https://doi.org/10.1080/0969594X.2019.1619515>.

- CHENG, L., ROGERS, T. and HU, H. (2004). “ESL/EFL Instructors’ Classroom Assessment Practices: Purposes, Methods, and Procedures”, **Language Testing**, Volume 21, Issue 3, Pages 360-389.
- DOI: <https://doi.org/10.1191/0265532204lt288oa>.
- CHOU, C. Y., LAI, K. R., CHAO, P. Y., LAN, C. H., and CHEN, T. H. (2015). “Negotiation Based Adaptive Learning Sequences: Combining Adaptivity and Adaptability”, **Computers & Education**, Volume 88, Issue 1, Pages 215-226.
- DOI: <https://doi.org/10.1016/j.compedu.2015.05.007>
- COOLEY, W. W. (1991). “State-Wide Student Assessment”, **Educational Measurement: Issues and Practice**, Volume 10, Pages 3-6.
- DOI: <https://doi.org/10.1111/j.1745-3992.1991.tb00209.x>.
- CORRERIA, R. C. (2016). “Assessing Speaking Proficiency: A Challenge for the Portuguese EFL Teacher”, **E-TEALS: E-Journal of Teacher Education and Applied Language Studies**, Volume 7, Issue 1, Pages 87-108. DOI: <https://doi.org/10.1515/eteals-2016-0009>.
- COWIE, B., and BELL, B. (1999). “A Model of Formative Assessment in Science Education”, **Assessment in Education**, Volume 6, Issue 1, Pages 101-116. DOI: <http://dx.doi.org/10.1080/09695949993026>.
- CRICHTON, H., and MCDAID, A. (2016). “Learning Intentions and Success Criteria: Learners’ and Teachers’ Views”, **The Curriculum Journal**, Volume 27, Issue 2, Pages 190-203. DOI: <https://doi.org/10.1080/09585176.2015.1103278>.
- CRISIANITA, S., and MANDASARI, B. (2022). “The Use of Small-Group Discussion to Improve Students’ Speaking Skill”, **Journal of English Language Teaching and Learning**, Volume 3, Issue 1, Pages 61-66.
- ÇETİN, Y. (2011). “Reliability of Raters For Writing Assessment: Analytic - Holistic, Analytic - Analytic, Holistic – Holistic”, **Mustafa Kemal University Journal of Social Sciences Institute**, Volume 8, Issue 16, Pages 471-486.

- DANG, G. (2017). "Teachers' Perceptions of Classroom-Based Assessment", **Journal of Foreign Language Studies**, Volume 53, Issue 4, Pages 74-81.
- DİKLİ, S. (2003). "Assessment at a Distance: Traditional vs. Alternative Assessments", **Turkish Online Journal of Educational Technology – TOJET**, Volume 2, Issue 3, Pages 13-19.
DOI: <https://doi.org/10.13140/RG.2.2.20943.12960>.
- DUNN, K. E., and MULVENON, S. W. (2009). "A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education", **Practical Assessment, Research and Evaluation**, Volume 14, Issue 7, Pages 1-11.
- EZİR, E. (2013). "Assessment and Testing in ELT: The Difference Between Assessment and Testing", **Journal SAINTIKOM**, Volume 12, Issue 1, Pages 37-42.
- ESTAJI, M, and KARDOUST, A. (2021). "Teachers' Use of Classroom-Based Assessment Strategies: Assessment Purposes and Practices", **Electronic Journal of Foreign Language Teaching [e-FLT]**, Volume 18, Issue 2, Pages 171-183. DOI: <https://doi.org/10.56040/etek1825>.
- FALCHIKOV, N. (2004). "Involving Students in Assessment", **Psychology Learning and Teaching**, Volume 3, Issue 2, Pages 102-108.
DOI: <https://doi.org/10.2304/plat.2003.3.2.1>.
- FARABI, M., HASSANVAND, S., and GORJIAN, B. (2017). "Using Guided Oral Presentation in Teaching English Language Learners' Speaking Skills", **Journal of Applied Linguistics and Language Learning**, Volume 3, Issue 1, Pages 17-24. DOI: <https://doi.org/10.5923/j.jalll.20170301.03>
- FARHADY, H. (1979). "The Disjunctive Fallacy between Discrete-Point and Integrative Tests", **TESOL Quarterly**, Volume 13, Issue 3, Pages 64-72.
DOI: <https://doi.org/10.2307/3585882>.
- FARHADY, H., and TAVASSOLI, K. (2021). "EFL Teachers' Perceptions and Practices of Their Language Assessment Knowledge", **Lang Test Asia**, Volume 11, Issue 17, Pages 1-19.

DOI: <https://doi.org/10.1186/s40468-021-00135-4>.

FAWZI, H. and ALDDABOUS, S. (2019). "Pre-Service Teachers' Perceptions and Preferences of Feedback Process", **International Journal of Education, Learning and Development**, Volume 7, Issue 1, Pages 36-47.

FRAIDAN, A. and MULHIM, N. (2012). "Direct and Indirect Assessments of Listening: An Investigation of the Nature of Short-Answer Questions and Dictation as Different Types of Testing Items", **International Journal of Linguistics**, Volume 4, Issue 3, Pages 149-178.

DOI: <https://doi.org/10.5296/ijl.v4i3.2200>.

FREY, N. and DOUGLAS, F. (2013). "A Formative Assessment System for Writing Improvement", **English Journal**, Volume 103, Issue 1, Pages 66-71.

FREY, B. and SCHMITT, V. L. (2010). "Teachers' Classroom Assessment Practices", **Middle Grades Research Journal**, Volume 5, Issue 3, Pages 107-117.

GHALIB, T. K., and AL-HATTAMI, A. A. (2015). "Holistic Versus Analytic Evaluation of EFL Writing: A Case Study", **English Language Teaching**, Volume 8, Issue 7, Pages 225-236.

DOI: <https://doi.org/10.5539/elt.v8n7p225>.

GIBBS, G AND SIMPSON, C., "Conditions Under Which Assessment Supports Students' Learning", **Learning and Teaching in Higher Education**, Volume 1, Issue 1, Pages 3-31.

GIBSON, S (2008). "Reading Aloud: A Useful Learning Tool?", **ELT Journal**, Volume 62, Issue 1, Pages 29-36.

DOI: <https://doi.org/10.1093/elt/ccm075>.

GIGANTE, J., DELL, M., and SHARKEY, A. (2011). "Getting Beyond "Good Job": How to Give Effective Feedback", **Pediatrics**, Volume 127, Issue 2, Pages 205-207. DOI: <https://doi.org/10.1542/peds.2010-3351>.

GILAKJANI, A. P., and SABOURI, N. B. (2017). "Teachers' Beliefs in English Language Teaching and Learning: A Review of the Literature", **English Language Teaching**, Volume 10, Issue 4, Pages 78-86.

DOI: <https://doi.org/10.5539/elt.v10n4p78>.

GÖKTÜRK, N. (2016). “Examining the Effectiveness of Digital Video recordings on Oral Performance of EFL Learners”, **Teaching English with Technology**, Volume 16, Issue 2, Pages 71-96.

GÖNEN, K. and AKBAROV, A. (2017).” Instructors’ Principles and Practices in the Implementation Stage of Classroom-Based Language Assessment in Higher Education in Turkey”, **Journal of European Education**, Volume 5, Issue 3, Pages 35-45.

GREEN, A. (2013). “Washback in Language Assessment”, **International Journal of English Studies**, Volume13, Issue 2, Pages 39–51.

DOI: <https://doi.org/10.6018/ijes.13.2.18589>.

HALADYNA, T. M., DOWNING, S. M., and RODRIGUEZ, M. C. (2002). “A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment”, **Applied Measurement in Education**, Volume 15, Issue 3, Pages 309–333.

DOI: https://doi.org/10.1207/S15324818AME1503_5.

HAMBLETON R. K., SWAMINATHAN H., ALGINA J., and COULSON D. B. (1978). “Criterion-Referenced Testing and Measurement: A Review Of Technical Issues and Developments”, **Review of Educational Research**, Volume 48, Issue 1, Pages 1–47.

DOI: <https://doi.org/10.3102/00346543048001001>.

HAMIDUN, N; HASHIM, S. H. M. and OTHMAN, N. F. (2012). “Enhancing Students’ Motivation by Providing Feedback on Writing: The Case of International Students from Thailand”, **International Journal of Social Science and Humanity**, Volume 2, Issue 6, Pages 591-594.

DOI: <https://doi.org/10.7763/IJSSH.2012.V2.179>.

HAMOUDA, A. (2013). “An Investigation of Listening Comprehension Problems Encountered by Saudi Students in the EL Listening Classroom”,., **International Journal of Academic Research in Progressive Education and Development**, Volume 2, Issue 2, Pages 113-15.

- HARDEN, R. M. (2002). "Developments in Outcome-Based Education", **Medical Teacher**, Volume 24, Issue 2, Pages 117-120.
DOI: <https://doi.org/10.1080/01421590220120669>.
- HARGREAVES, E (2005). "Assessment for Learning? Thinking Outside the (Black) Box", **Cambridge Journal of Education**, Volume 35, Issue 2, Pages 213-224.
- HARLAN, W., and WINTER, J. (2004). "The Development of Assessment for Learning: Learning from The Case of Science and Mathematics", **Language Testing**, Volume 21, Issue 3, Pages 390-408.
DOI: <https://doi.org/10.1191/0265532204lt289oa>.
- HASROL, S. B., ZAKARIA, A., and ARYADOUST, V. (2022). "A Systematic Review of Authenticity in Second Language Assessment", **Research Methods in Applied Linguistics**, Volume 1, Issue 3, Pages 100023.
DOI: <https://doi.org/10.1016/j.rmal.2022.100023>.
- HATZIAPOSTOLOU, T., and PARASKAKIS, I. (2010). "Enhancing the Impact of Formative Feedback on Student Learning Through an Online Feedback System", **Electronic Journal of E-learning**, Volume 8, Issue 2, Pages 111-122.
- HAVNES, A., SMITH, K., DYSTHE, O., and LUDVIGSEN, K. (2012). "Formative Assessment and Feedback: Making Learning Visible", **Studies in Educational Evaluation**, Volume 38, Issue 1, Pages 21-27.
DOI: <https://doi.org/10.1016/j.stueduc.2012.04.001>.
- HAYATI, A., and JALILIFAR, A. (2009). "The Impact of Note-Taking Strategies on Listening Comprehension of EFL Learners". **English Language Teaching**, Volume 2, Issue 1, Pages 101-111.
DOI: <https://doi.org/10.5539/elt.v2n1p101>.
- HEMMATI, F., and GHADERI, E. (2014). "The Effect of Four Formats of Multiple-Choice Questions on the Listening Comprehension of EFL Learners", **Procedia - Social and Behavioral Sciences**, Volume 98, Issue 1, Pages 637-644.

DOI: <https://doi.org/10.1016/j.sbspro.2014.03.462>

HIGGINS, R., HARTLEY P., and SKELTON A. (2001). “Getting the Message Across: The Problem of Communicating Assessment Feedback”, **Teaching in Higher Education**, Volume 6, Issue 2, Pages 269–74.

DOI: <https://doi.org/10.1080/13562510120045230>.

HECK, R. H., and CRISLIP, M. (2001). “Direct and Indirect Writing Assessments: Examining Issues of Equity and Utility”. **Educational Evaluation and Policy Analysis**, Volume 23, Issue1, Pages 19–36 DOI:

<https://doi.org/10.3102/01623737023001019>.

HIRAI, A., and KOIZUMI, R. (2009). “Development of a Practical Speaking Test with A Positive Impact on Learning Using A Story Retelling Technique”, **Language Assessment Quarterly**, Volume 6, Issue 2, Pages 151-167.

DOI: <https://doi.org/10.1080/15434300902801925>

HOVARDAS, T., TSIVITANIDOU, O. E., and ZACHARIA, Z. C. (2014). “Peer Versus Expert Feedback: An Investigation of the Quality of Peer Feedback Among Secondary School Students”, **Computers and Education**, Volume 71, Issue 1, Pages 133–152.

DOI: <https://doi.org/10.1016/j.compedu.2013.09.019>.

HUANG, L. (2010). “Reading Aloud in The Foreign Language Teaching”, **Asian Social Science**, Volume 6, Issue 4, Pages 148-150.

DOI: <https://doi.org/10.5539/ass.v6n4p148>.

HUDDLESTON, E. M. (1954). “Measurement of Writing Ability at the College Entrance Level: Objective vs. Subjective Testing Techniques”, **The Journal of Experimental Education**, Volume 22, Issue 3, Pages 165-213.

HUERTA-MACIAS, A. (1995). “Alternative Assessment: Responses to Commonly Asked Question”. **TESOL Journal**, Volume 5, Issue 1, Pages 8-11.

IDHAM, S. Y., BAAGBAH, S. Y. S., MUGAIR, S. K., FENG, H., and AL HUSSEINY, F. (2024). “Impact of the Interviews on the Students’

- Speaking Skills”, **Theory and Practice in Language Studies**, Volume 14, Issue 1, Pages 160-166. DOI: <https://doi.org/10.17507/tpsl.1401.19>.
- IMRAN, K. (2011). “Reading Assessment Techniques Among Selected Secondary School Teachers in Pakistan: Current Trends and Practices”, **International Journal on New Trends in Education and their Implications**, Volume 2, Issue 4, Pages 58-141.
- JANNATI, S. (2015). “ELT Teachers’ Language Assessment Literacy: Perceptions and Practices”, **Educational Research Association the International Journal of Research in Teacher Education**, Volume 6, Issue 2, Pages 26-37.
- JÖNSSON, A., BALAN, A., and HARTELL, E. (2021). “Analytic or Holistic? A study About How to Increase the Agreement in Teachers’ Grading”, **Assessment in Education: Principles, Policy and Practice**, Volume 28, Issue 3, Pages 212–227.
DOI: <https://doi.org/10.1080/0969594x.2021.1884041>.
- KAÇAR, K. (2023). “The purpose and Principles of Testing in ELT”, **Alanyazın**, Volume 4, Issue 1, Pages 33-41.
DOI: <https://doi.org/10.59320/alanyazin.1255021>.
- KARİM, S., and HAQ, N. (2014). “An Assessment of IELTS Speaking Test”, **International Journal of Evaluation and Research in Education**, Volume 3, Issue 3, Pages 152-157,
DOI: <https://doi.org/10.11591/ijere.v3i3.6047>.
- KASSEM, H. (2015). “The Relationship between Listening Strategies Used by Egyptian EFL College Sophomores and Their Listening Comprehension and Self-Efficacy”, **English Language Teaching**, Volume 8, Issue 2, Pages 153-169. DOI: <http://dx.doi.org/10.5539/elt.v8n2p153>.
- KETABI, S. (2016). “Methods of Assessing Reading Used by Iranian EFL Teachers”, **International Journal of Research Studies in Education**, Volume 5, Issue 2, Pages 29-42.

- KHODADADY, E., and KHODABAKHSHZADE, H. (2012). "The Effect of Portfolio and Self Assessment on Writing Ability and Autonomy". **Journal of Language Teaching & Research**, Volume 3, Issue 3, Pages 518-524. DOI: <https://doi.org/10.4304/jltr.3.3.518-524>.
- KIRKGÖZ, Y., BALABANOĞLU, M. P. and AGCAM, R. (2017). "Turkish EFL Teachers' Perceptions and Practices of Foreign Language Assessment in Primary Education", **Journal of Education and e-Learning Research**, Volume 4, Issue, 4, Pages 163-170.
DOI: <https://doi.org/10.20448/journal.509.2017.44.163.170>.
- KING, J. (2002). "Preparing EFL Learners for Oral Presentations" **The Internet TESL Journal**, Volume 8, Issue 3, Pages 401-417.
- KINTSCH, W., and VAN DIJK, T. A. (1978). "Toward a Model of Text Comprehension and Production", **Psychological Review**, Volume 85, Issue 5, Pages 363-394.
- KLENOWSKI, V. (2009). "Assessment for Learning Revisited: An Asia-Pacific Perspective", **Assessment in Education: Principles, Policy and Practice**, Volume 16, Issue 3, Pages 263–268.
DOI: <https://doi.org/10.1080/09695940903319646>.
- KLINGNER, J. (2004). "Assessing Reading Comprehension", **Assessment for Effective Intervention**, Volume 29, Issue 4, Pages 59-70.
DOI: <https://doi.org/10.1177/073724770402900408>.
- KORKMAZ, Ş. Ç., (2021). "Target Language Use in EFL Classrooms: Turkish EFL Teachers' Beliefs and Practices", **Indonesian Journal of English Language Teaching and Applied Linguistics**, Volume 6, Issue 1, Pages 155-171.
- KUMPAS-LENK, K., EISENSCHMIDT, E., and VEISPAK, A. (2018). "Does the design of learning outcomes matter from students' perspective?", **Studies in Educational Evaluation**, Volume 59, Issue 1, Pages 179-186.
DOI: <https://doi.org/10.1016/j.stueduc.2018.07.008>.

- KÜRÜM, E. Y. (2016). "Teaching Speaking Skills", **Annals of the University of Oradea: Economic Science**, Volume 25, Issue 1, Pages 264-273.
- 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**, Volume 41, Issue 18, Pages 1-18.
DOI: <http://dx.doi.org/10.1080/03075079.2014.999317>.
- LATIF, M. W., and WASIM, A. (2022). "Teacher Beliefs, Personal Theories and Conceptions of Assessment Literacy—A Tertiary EFL Perspective", **Language Testing in Asia**, Volume 12, Issue 1, Page 1-22.
DOI: <https://doi.org/10.1186/s40468-022-00158-5>.
- LEE, I. (2007). "Feedback in Hong Kong Secondary Writing Classrooms: Assessment for Learning or Assessment of Learning?", **Assessing Writing**, Volume 12, Issue 3, Pages 180–198.
DOI: <https://doi.org/10.1016/j.asw.2008.02.003>.
- LEECK, P. (2015). "Portfolio-Work and Its Effects on Listening Comprehension of Very Young Learners", **Journal of Language and Cultural Education**, Volume 3, Issue 1, Pages 43–56.
DOI: <https://doi.org/10.1515/jolace-2015-0004>.
- LI, X., (2012). "The Role of Teachers' Beliefs in the Language Teaching-Learning Process", **Theory and Practice**, Volume 2, Issue 7, Pages 1397-1402.
DOI: <https://doi.org/10.4304/tpls.2.7.1397-1402>.
- LI, W. (2011). "Validity Considerations in Designing an Oral Test", **Journal of Language Teaching and Research**, Volume 2, Issue 1, Pages 267-269.
DOI: <https://doi.org/10.4304/jltr.2.1.267-269>.
- LIN, Y. and LIN C. (2019). "Understanding How Teachers Practise Inclusive Classroom Assessment", **Studies in Educational Evaluation**, Volume 63, Issue 4, Pages 113-121.
- LIU, F., and DING, Y. (2009). "Role-Play in English Language Teaching", **Asian Social Science**, Volume 5, Issue 10, Pages 140-143.

DOI: <https://doi.org/10.5539/ass.v5n10p140>.

LIU, N. F., and CARLESS, D. (2006). "Peer Feedback: The Learning Element of Peer Assessment", **Teaching in Higher education**, Volume 11, Issue 3, Pages 279-290. DOI: <https://doi.org/10.1016/j.compedu.2013.09.019>.

MAK, B. (2011). "An Exploration of Speaking-In-Class Anxiety with Chinese ESL Learners", **System**, Volume 39, Issue 2, Pages 202-214. DOI: <https://doi.org/10.1016/j.system.2011.04.002>.

MAVROMMATIS, Y. (1997). "Understanding Assessment in the Classroom: Phases of the Assessment Process—The Assessment Episode", **Assessment in Education: Principles, Policy and Practice**, Volume 4, Issue 3, Pages 381-400. DOI: <https://doi.org/10.1080/0969594970040305>.

MCFADZIEN, N. (2015). "Why is Effective Feedback So Critical in Teaching and Learning?", **Journal of Initial Teacher Inquiry**, Volume 1, Issue 1, Pages 16- 18.

MCINTOSH, J. E. (2010). "Reader Response Journals: Novice Teachers Reflect on Their Implementation Process", **Journal of Language and Literacy Education**, Volume 6, Issue 1, Pages 119-133.

MEDE, E. and ATAY, D. (2017). "English Language Teachers' Assessment Literacy: the Turkish Context", **Dil Dergisi**, Volume 168, Issue 1, Pages 1-5.

MEKONNEN, G. T. (2014). "EFL Classroom Assessment: Teachers Practice and Teaching Techniques Adjustment in Ethiopia", **Educational Research and Reviews**, Volume 9, Issue 20, Pages 1071-1089.

DOI: <https://doi.org/10.5897/ERR2014.1857>.

MIRMAKHMUDOVA, I. I. (2021). "Comparing Criterion and Norm Referenced Assessments of Language Skills in the Second Language", **Asian Journal of Research in Social Sciences and Humanities**, Volume 11, Issue 11, Pages 463–468. DOI: <https://doi.org/10.5958/2249-7315.2021.00242.2>.

MULVENON, S.W. (2009). "A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence on the Impact of

Assessment in Education. **Practical Assessment, Research, and Evaluation**, Volume 14, Issue 1, Pages 1-11.

DOI: <https://doi.org/10.7275/jg4h-rb87>.

MUNIF, A.L., FAUZIATI, E., and MARMANDO, S. (2019). “Performance Assessment: Teachers’ Beliefs and Practices in Higher Education”, **Journal of Education and Learning**, Volume 13, Issue 4, Pages 518-526. DOI: <https://doi.org/10.11591/edulearn.v13i4.13068>.

MUÑOZ, A. P., PALACIO, M., and ESCOBAR, L. (2012). “Teachers’ Beliefs About Assessment in an EFL Context in Colombia”, **Profile Issues in Teachers Professional Development**, Volume 14, Issue 1, Pages 143-158.

NICOL, D. (2019). “Reconceptualising Feedback as an Internal Not an External Process”, **Italian Journal of Educational Research**, Special Issue, Pages 71–83.

NDOYE, A. (2017). “Peer/Self-assessment and Student Learning”, **International Journal of Teaching and Learning in Higher Education**, Volume 29, Issue 2, Pages 255-269.

NURHAYATI, N., BAIQ, S. H., and ZAH RATUL, F. (2020). “An Analysis of Teachers’ and Students’ Perceptions on Formal and Informal Assessment”, **Journal on Language and Literature**, Volume 6, Issue 2, Pages 143-150.

O’LOUGHLIN, K. (2006). “Learning About Second Language Assessment: Insights from A Postgraduate Student On-Line Subject Forum”, **University of Sydney Papers in TESOL**, Volume 1, Pages 71-85.

OMELICHEVA, M. Y. (2005). “Self and Peer Evaluation in Undergraduate Education: Structuring Conditions That Maximize Its Promises and Minimize the Perils”, **Journal of Political Science Education**, Volume 1, Issue 2, Pages 191-205.

DOI: <https://doi.org/10.1080/15512160590961784>.

- OPRE, D. (2015). "Teachers' Conceptions of Assessment", **Procedia-Social and Behavioral Sciences**, Volume 209, Issue 1, Pages 229-233, DOI: <https://doi.org/10.1016/j.sbspro.2015.11.222>.
- ÖZ, H. (2014). "Turkish Teachers' Practices of Assessment for Learning in the English as a Foreign Language Classroom", **Journal of Language Teaching and Research**, Volume 5, Issue 4, Pages 775-785.
DOI: <https://doi.org/10.4304/jltr.5.4.775-785>.
- ÖZ, S. and ATAY, D. (2017). "Turkish EFL Instructors in-Class Language Assessment Literacy: Perceptions and Practices", **ELT Research Journal**, Volume 6, Issue 1, Pages 25-44.
- ÖLMEZLER ÖZTÜRK, E. and AYDIN, B. (2019). "Investigating Language Assessment Knowledge of EFL Teachers", **Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education)**, Volume, 34, Issue 3, Pages 602-620.
DOI: <https://doi.org/10.16986/HUJE.2018043465>.
- ÖNALAN, O., and GÜRSOY, E. (2020). "Private School EFL Teachers' Perceptions on Assessment: A Case Study in Turkey", **Porta Linguarum**, Volume 2020, Issue 33, Pages 249-264.
- ÖNALAN, O., and 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**, Volume 14, Issue 3, Pages 190-201.
- ÖZDEMİR YILMAZER, O. and ÖZKAN, Y. (2017). "Classroom Assessment Practices of English Language Instructors", **Journal of Language and Linguistic Studies**, Volume 13, Issue, 2, Pages 324-345.
- PAILY, M. U. (2013). "Creating Constructivist Learning Environment: Role of "Web 2.0" Technology", **International Forum of Teaching and Studies**, Volume 9, Issue 1, Pages 39-50.
- PAXTON, M. (2000). "A Linguistic Perspective on Multiple Choice Questioning", **Assessment & Evaluation in Higher Education**, Volume 25, Issue 2, Pages 109-119. DOI: <https://doi.org/10.1080/713611429>.

- PIPER, K., MORPHET, J. and BONNAMY, J. (2019). "Improving Student-Centered Feedback Through Self-assessment", **Nurse Educ Today**, Volume 83, Issue 1, Pages 1-6. DOI: <https://doi.org/10.1016/j.nedt.2019.08.011>.
- PHONGSIRIKUL, M. (2018), "Traditional and Alternative Assessments in ELT: Students' and Teachers' Perceptions", **REFlections**, Volume 25, Issue 1, Pages 61-84. DOI: <https://doi.org/10.61508/refl.v25i1.136267>.
- POLE, KATHRYN (2007) "Mixed Method Designs: A Review of Strategies for Blending Quantitative and Qualitative Methodologies", **Mid-Western Educational Researcher**, Vol. 20, Issue 4, Pages 35-38.
- PONTO, J. (2015). "Understanding and Evaluating Survey Research", *Journal of the Advanced Practitioner in Oncology*, Volume 6, Issue 2, Pages 168-171.
- PURNAMAWATI, S., and SUHARTONO, L. (2015). "The Use of Role Play In Teaching Speaking", **Jurnal Pendidikan Dan Pembelajaran Khatulistiwa (JPPK)**, Volume 4, Issue 9, Pages 1-13.
- QIAN, D. (2009). "Comparing Direct and Semi-Direct Modes for Speaking Assessment: Affective Effects on Test Takers", **Language Assessment Quarterly**, Volume 6, Issue 2, Pages 113-125,
DOI: <https://doi.org/10.1080/15434300902800059>.
- QUANSAH, F. (2018). "Traditional or Performance Assessment: What Is the Right Way in Assessing Learners?" **Research on Humanities and Social Sciences**, Volume 8, Issue 1, Pages 21-24.
- QUILTER, S. M., and GALLINI, J. K. (2000). "Teachers' Assessment Literacy and Attitudes", **The Teacher Educator**, Volume 36, Issue 2, Pages 115–131.
- RAO, S. P. (2019). "The Importance of Speaking Skills in English Classrooms", **Alford Council of International English and Literature Journal**, Volume 2, Issue, 2, Pages 6-18.
- RAHMATI, T., and NUSHI, M. (2023). "EFL Teachers' Knowledge, Beliefs, and Practices Regarding Fairness and Justice in Technology-Enhanced Classroom Assessment: A Duoethnography". **Iranian Journal of Language Teaching Research**, Volume 11, Issue 3, Pages 31-52.

DOI: <https://doi.org/10.30466/ijltr.2023.121405>.

RASHID, S., MAHMOOD, N. (2020). "High Stake Testing: Factors Affecting Inter-Rater Reliability in Scoring of Secondary School Examination", **Bulletin of Education and Research**, Volume 42, Issue 2, Pages 163-179.

REA-DICKINS, P. (2004). "Understanding Teachers as Agents of Assessment", **Language Testing**, Volume 21, Issue 3, Pages 249-258.

DOI: <https://doi.org/10.1191/0265532204lt283ed>.

ROLLINSON, P. (2005). "Using Peer Feedback in the ESL Writing Class", **ELT Journal**, Volume 59, Issue 1, Pages 23-30,

DOI: <https://doi.org/10.1093/elt/cci003>.

RUKTHONG, A. (2021). "MC Listening Questions Vs. Integrated Listening-To-Summarize Tasks: What Listening Abilities Do They Assess?", **System**, Volume 97, Issue 1, Pages 1-12.

DOI: <https://doi.org/10.1016/j.system.2020.102439>.

RUKTHONG, A., and BRUNFAUT, T. (2019). "Is Anybody Listening? The Nature of Second Language Listening in Integrated Listening-To-Summarize Tasks", **Language Testing**, Volume 37, Issue 1, Pages 31–53.

DOI: <https://doi.org/10.1177/0265532219871470>.

RUSSO, C. (2020). "The Impacts of Individual Reading Conferences on Student Growth and Reading Confidence", **Florida Journal of Educational Research**, Volume 58, Issue 7, Pages 93–101.

SAAD, M. R. B. M., SARDAREH, S. A., and AMBARWATI, E. K. (2013). "Iranian Secondary School EFL Teachers' Assessment Beliefs and Roles". **Life Science Journal**, Volume 10, Issue 3, Pages 1638-1647.

SADLER, D.R. (1989). "Formative Assessment and the Design of Instructional System", **Instructional Science**, Volume 18, Issue 2, Pages 119-144.
DOI: <https://doi.org/10.1007/BF00117714>.

SAEFURROHMAN, PH.D. (2018). "EFL Teachers Assessment Methods in Oral Communications", **Advances in Social Science, Education and Humanities Research (ASSEHR)**, Volume 267, Issue 1, Pages 268-272.

- SAFEI, N. H. (2020). "The Effectiveness of Retelling Short Story Towards Students' Accuracy in Speaking Skill", **IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature**, Volume 8, Issue 1, Pages 137-144. DOI: <https://doi.org/10.24256/ideas.v8i1.1337>.
- SHABAN, A. (2014). "Comparison Between Objective and Subjective Tests", **Journal of the College of Languages**, Volume 1, Issue 30, Pages 44-52.
- SCHULZ, R. A. (1977). "Discrete-Point versus Simulated Communication Testing in Foreign Languages", **The Modern Language Journal**, Volume 61, Issue 3, Pages 94–101. DOI: <https://doi.org/10.2307/325537>.
- SHUTE, V. J. (2008). "Focus on Formative Feedback", **Review of Educational Research**, Volume 78, Issue 1, Pages 153-189.
DOI: <https://doi.org/10.3102/0034654307313795>.
- SİDDİEK, A. G. (2010). "The Impact of Test Content Validity on Language Teaching and Learning", **Asian Social Science**, Volume 6, Issue 12, Pages 133-143.
DOI: <https://doi.org/10.5539/ass.v6n12p133>.
- SPINELLI, C. G. (2000). "Diagnostic and Prescriptive Assessment Strategies for the General and Special Education Teacher", **Hong Kong Special Education Forum**, Volume 3, Issue 1, Pages 1–21.
- STIGGINS, R. (2007), "Assessment Through the Student's Eyes", **Educational Leadership: Journal of the Department of Supervision and Curriculum Development**, Volume 64, Issue 8, Pages 22-26.
- STIGGINS R., and CHAPPUS J. (2005). "Using Student-Involved Classroom Assessment to Close Achievement Gaps", **Theory into Practice**, Volume 44, Issue 1, Pages 11-18.
DOI: https://doi.org/10.1207/s15430421tip4401_3.
- SUSILO, R., MARBUN, R., and HUSIN, S. (2016). "The Use of Retelling Story to Improve Students' Ability in Speaking", **Jurnal Pendidikan dan Pembelajaran Khatulistiwa (JPPK)**, Volume 5, Issue 3, Pages 1-8.
- SYAKUR, A. (2020). "Improving the Eighth Grade Students' Listening Comprehension Achievement by using Dictation Techniques",

Konfrontasi: Jurnal Kultural, Ekonomi Dan Perubahan Sosial,
Volume 7, Issue 3, Page 205-216.

DOI: <https://doi.org/10.33258/konfrontasi2.v7i3.116>.

SWAFFIELD, S. (2011). "Getting to The Heart of Authentic Assessment for Learning", **Assessment in Education: Principles, Policy and Practice**, Volume 18, Issue 4, Pages 433-449.

DOI: <https://doi.org/10.1080/0969594X.2011.582838>.

TAHERDOOST, H. (2016). "Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research", **International Journal of Academic Research in Management (IJARM)**, Volume 5, Issue 3, Pages 28-36.

TAJIK, L. and RANJBAR, K. (2018). "Reflective Teaching In ELT: Obstacles and Coping Strategies", **Journal of Research in Applied Linguistics**, Volume 9, Issue 1, Pages 148-169.

DOI: <https://doi.org/10.22055/rals.2018.13409>.

TANNER, D. E. (2001). "Authentic Assessment: A Solution, or Part of the Problem?" **High School Journal**, Volume 85, Issue 1, Pages 24-29.

DOI: <https://doi.org/10.2139/ssrn.3205040>.

TAVİL, Z. M. (2014). "Integrating Listening and Speaking Skills to Facilitate English Language Learners' Communicative Competence", **Procedia - Social and Behavioral Sciences**, Volume 9, Issue 1, Pages 765-770.

DOI: <https://doi.org/10.1016/j.sbspro.2010.12.231>.

THOMAS, M. (2012). "Teachers' Beliefs About Classroom Assessment and Their Selection of Classroom Assessment Strategies", **Journal of Research and Reflections in Education (JRRE)**, Volume 6, Issue 2, Pages 103-112.

DOI: <https://doi.org/10.1016/j.sbspro.2013.08.805>.

TIERNEY, R. D. (2006). "Changing Practices: Influences on Classroom Assessment", **Assessment in Education**, Volume 13, Issue 3, Pages 239-264.

DOI: <https://doi.org/10.1080/09695940601035387>.

TOSUNCUOĞLU, I. (2018). “Importance of Assessment in ELT”. **Journal of Education and Training Studies**, Volume 6, Issue 9, Pages 163-167.

ÜLKER, V., YILDIZ, Y. and BARZANI, T. F. (2021). “The Implementation of the Authentic Assessment in EFL Classes in Erbil”, **International Journal of Social Sciences & Educational Studies**, Volume 8, Issue 2, Pages 206-221. DOI: <https://doi.org/10.23918/ijsses.v8i2p206>.

VAN HA, X., TRAN, N. G., and TRAN, N. H. (2021). “Teachers’ Beliefs and Practices Regarding Assessment in English as A Foreign Language Classrooms in Vietnam”. **The Qualitative Report**, Volume 26, Issue 11, Pages 3457-3475. DOI: <https://doi.org/10.46743/2160-3715/2021.5063>.

VICKERMAN, P. (2009). “Student Perspectives on Formative Peer Assessment: An Attempt to Deepen Learning”, **Assessment and Evaluation in Higher Education**, Volume 34, Issue 2, Pages 221-230.

DOI: <https://doi.org/10.1080/02602930801955986>.

WANG, L., LEE, I., and PARK, M. (2020). “Chinese University EFL Teachers’ Beliefs and Practices of Classroom Writing Assessment”. **Studies in Educational Evaluation**, Volume 66, Issue 1, Pages 1-11.

DOI: <https://doi.org/10.1016/j.stueduc.2020.100890>.

WIDIASTUTI, I. A. M. S., MUKMINATIEN, N., PRAYOGO, J. A., and IRAWATI, E. (2020). “Dissonances between Teachers’ Beliefs and Practices of Formative Assessment in EFL Classes”, **International Journal of Instruction**, Volume 13, Issue 1, Pages 71-84.

DOI: <https://doi.org/10.29333/iji.2020.1315a>.

WIGGINS, G. (2016). “Seven Keys to Effective Feedback”, **Educational Leadership: Journal of the Department of Supervision and Curriculum Development**, Volume 70, Issue 1, Pages 10-16.

WILLIS, J., ARNOLD, J., and DELUCA, C. (2023,). Accessibility, in Assessment for Learning: Sharing Criteria for Success”, **Frontiers in Education**,

Volume 8, Issue 1, Pages 1-15. DOI:
<https://doi.org/10.3389/feduc.2023.1170454>.

WOLF, K., and SIU-RUNYAN, Y. (1996). "Portfolio Purposes and Possibilities", **Journal of Adolescent and Adult Literacy**, Volume 40, Issue 1, Pages 30–37.

WOLF, M. K., FRANSWORTH, T. and HERMAN, J. L. (2008). "Validity Issues in Assessing English Language Learners' Language Proficiency", **Educational Assessment**, Volume 13, Issue 2, Pages 80-107.

DOI: <http://www.doi.org/10.1080/10627190802394222>.

YAN, Z., and BROWN, G. T. L. (2017). "A Cyclical Self-Assessment Process: Towards A Model of How Students Engage in Self-Assessment". **Assessment & Evaluation in Higher Education**, Volume 42, Issue 8, Pages 1247–1262. DOI:

<https://doi.org/10.1080/02602938.2016.1260091>.

YAN, Z., WANG, X., BOUD, D., and LAO, H. (2023). "The Effect of Self-Assessment on Academic Performance and the Role of Explicitness: A Meta-Analysis", **Assessment and Evaluation in Higher Education**, Volume 48, Issue 1, Pages 1-15.

DOI: <https://doi.org/10.1080/02602938.2021.2012644>.

YANG, L., and WILSON, K. (2006). "Second Language Classroom Reading: A Social Constructivist Approach", **The Reading Matrix**, Volume 6, Issue 3, Pages 364-372.

YAO, Y. (2015). "Teacher Perceptions of Classroom Assessment: A Focus Group Interview", **SRATE Journal**, Volume 24, Issue 2, Pages 51-58.

YURDABAKAN, İ. (2011). "The View of Constructivist Theory on Assessment: Alternative Assessment Methods in Education", **Ankara University Journal of Faculty of Educational Sciences (JFS)**, Volume 44, Issue 1, Pages 51-78. DOI: https://doi.org/10.1501/Egifak_0000001215.

ZAINUDDIN, M. (2004). "The Effect of Giving Feedback to Students' Writing", **TEFLIN Journal**, Volume 15, Issue 2, Pages 117-126.

- ZHANG, Z. and BURRY-STOCK, J. A. (2003). "Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills", **Applied Measurement in Education**, Volume 16, Issue 4, Pages 324-342. DOI: https://doi.org/10.1207/S15324818AME1604_4.
- ZHANG, B., XIAO, Y. and LUO, J. (2015). "Rater Reliability and Score Discrepancy Under Holistic and Analytic Scoring of Second Language Writing", **Language Testing in Asia**, Volume 5, Issue 1, Pages 1-9.
- ZOLFAGHARI, A. (2018). "EFL Teachers' Assessment Literacy and Their Reflective Teaching", **International Journal of Instruction**, Volume 11, Issue 1, Pages 425-436.
DOI: <https://doi.org/10.12973/iji.2018.11129a>.
- ZULAIHA, S., MULYONO, H., and AMBARSARI, L. (2020). "An Investigation into EFL Teachers' Assessment Literacy: Indonesian Teachers' Perceptions and Classroom Practice". **European Journal of Contemporary Education**, Volume 9, Issue 1, Pages 189-201.
DOI: <https://doi.org/10.1186/s40468-015-0014-4>.

ELECTRONIC SOURCES

- ADAM, S, "Using Learning Outcomes: A Consideration of the Nature, Role, Application and Implications for European Education of Employing "Learning Outcomes" at the Local, National and International Levels. United Kingdom Bologna Seminar, 1-2 July 2004, Heriot-Watt University (Edinburgh Conference Centre), Edinburgh.
www.qualityresearchinternational.com/glossary/learningoutcomes.htm.
(Access Date: 02 June 2024)
- ASSESSMENT REFORM GROUP, "Assessment for Learning: 10 Principles. University of Cambridge, Faculty of Education", University of Cambridge, Faculty of Education, 2002,
<https://www.storre.stir.ac.uk/bitstream/1893/32458/1/Assessment%20for%20learning%2010%20principles%202002.pdf>. (Access Date: 08 June 2024)

DEAKIN CRICK, RE., and HARLEN, W. “A Systematic Review of The Impact of Summative Assessment and Testing on Pupil's Motivation For Learning”, 2002, Research Evidence in Education Library.

<http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=108>. (Access Date: 15 May 2024)

EARL L. M., Katz, MANITOBA, M. S. and Western and Northern Canadian Protocol for Collaboration in Education, “Rethinking Classroom Assessment with Purpose in Mind: Assessment for Learning, Assessment as Learning Assessment of Learning”. Manitoba Education Citizenship & Youth, 2006,

http://www.edu.gov.mb.ca/ks4/assess/wncp/rethinking_assess_mb.pdf. (Access Date: 10 June 2024)

ET, O. D. C., “Assessment for Learning Formative Assessment”, 2008,

<https://www.oecd.org/site/educeri21st/40600533.pdf>. (Access Date: 10 June 2024)

GRAHAM, S., HARRIS, K., and HEBERT, M. A, “Informing Writing: The Benefits of Formative Assessment”. A Carnegie Corporation Time to Act Report, Washington, DC: Alliance for Excellent Education, 2011,

https://media.carnegie.org/filer_public/37/b8/37b87202-7138-4ff9-90c0-cd6c6f2335bf/ccny_report_2011_informing.pdf. (Access Date: 2 January 2024)

HERITAGE, M, “Formative Assessment: Making It Happen in the Classroom” 2010,

<https://dx.doi.org/10.4135/9781452219493>. (Access Date: 6 February 2024)

JACOB, C. L., “How to Write Better Tests: A Handbook for Improving Test Construction Skills”, 2010,

http://www.indiana.edu/~best/write_better_tests.shtml. (Access Date: 20 September 2023)

HATİPOĞLU, “Testing and Assessment of Speaking Skills, Test Task Types, and Sample Test Items”, 2021, <https://hdl.handle.net/11511/98974>. (Access Date: 20 September 2023).

KARMILA, S., and JUFRIZAL, R. "Teachers Practices on Speaking Assessment of English as a Foreign Language A Descriptive Study at Junior High Schools in Sintuk Toboh Gadang Subdistrict", In 1st International Conference on Education Social Sciences and Humanities, 2019, <https://doi.org/10.2991/icesshum-19.2019.39>. (Access Date: 20 May 2024)

WILLIAM, D., "Integrating Formative and Summative Functions of Assessment", 2000, <https://core.ac.uk/download/pdf/79527875.pdf>. (Access Date: 11 June 2024)

DISSERTATIONS

AYAŞ, N. (2014). "An Investigation of Teachers' Conception and Practices Concerning Assessment in English For Young Learners Classroom" (Unpublished MA Thesis), Department of English Language Teaching, Çağ University.

BALBOA ÁLVAREZ, C. (2020). "The Power of Peers: A Study on Teachers' Beliefs on Peer and Self-Assessment in The EFL Upper-Secondary School Classroom in Sweden" (Unpublished MA Thesis), Master of Arts/Science in Upper-Secondary Education, Malmö University.

BAŞAK, H. (2019). "Self-Assessment of Students' Speaking Skills" (Unpublished MA Thesis), Department of English Language Teaching, Middle East Technical University.

BELLE, D. (1999). "Traditional Assessment vs. Alternative Assessment" (Unpublished MA Thesis), Department of Education, Kean University.

BURAL, E. (2021). "EFL Teachers' Beliefs on the Applicability of Dynamic Assessment in Two Different Contexts: Turkey and Spain", (Unpublished MA Thesis), Department of English Language Teaching, Kocaeli University.

CAN, E. (2017). "English Teachers' Classroom Assessment Practices and Their Views About the Ethics of Classroom Assessment Practices", (Unpublished MA Thesis), Department of English Language Education, Çağ University.

- ÇAKIR, N. (2020). "A Comparative Analysis of Teachers' Beliefs and Practices on The Assessment of 4th Grade-EFL Students in Turkey, Italy And Finland" (Unpublished MA Thesis), Department of Foreign Language Education, Bursa Uludağ University.
- ERDEN BURNAZ Y. (2011). "Perceptions of EFL Learners Towards Portfolios as A Method of Alternative Assessment: A Case Study at A Turkish State University", (Unpublished MA Thesis), Department of English Language Teaching, Middle East Technical University.
- MKHWANAZI, H. N. (2014). "Teachers' Use of Formative Assessment in the Teaching of Reading Comprehension in Grade 3" (Unpubished Doctoral Dissertation), Department of Language Teaching, University of Pretoria.
- MUTLU, A (2022). "Turkish EFL Teachers' Beliefs, Practices and Challenges on Assessment" (Unpublished MA Thesis), Department of Foreign Language Education English Language Teaching Program, Pamukkale University.
- ÖRSDERİR, E. (2010). "Alternative Assessment in Turkish Primary EFL Classrooms: An investigation Into the Performance Tasks", (Unpublished MA Thesis), Department of English Language Education, Çukurova University.
- ÖZUSLU, N. (2018). "An Evaluation of the Performance Assessment Component of an EFL, Preparatory School Program", (Unpublished MA Thesis), Department of English Language Education, Yeditepe University.
- SWAI, M. T. A. (2019). "An Investigation of Jordanian English as a Foreign Language (EFL) Teachers' Perceptions of and Practices in Classroom Assessment", (Unpublished MA Thesis), Applied Linguistics and Discourse Studies, Carleton University.
- ŞENGÜL, M. (2022). "Assessment Practices in The Preparatory Language Schools in Turkish Higher Education Institutions: Problems and Solutions", (Unpublished PhD Thesis), Department of Foreign Language Education, Anadolu University.

- TÜMER, M. (2019). “A Study on Using Visual Elements in Assessing Listening Comprehension”, (Unpublished MA Thesis), Department of Foreign Language Teaching, Ufuk University.
- UZUN, G. (2007). “The Role of Classroom Assessment in Teaching English”, (Unpublished MA Thesis), Department of English Teaching, Selçuk University.
- YETKİN, C. (2015). “An Investigation on ELT Teacher Candidates’ Assessment Literacy” (Unpublished MA Thesis), Department of English Language Teaching, Çağ University.



APPENDICES

Appendix A: Questionnaire

Appendix B: Interview Questions

Appendix C: Ethics Committee Decision





Appendix A: Questionnaire

PART1: Background Information

*Gender: ☐ Male ☐ Female ☐ Other

*Age: ☐ 18-25 ☐ 26-30 ☐ 31-35 ☐ 36-40 ☐ 41-45 ☐ 46-50 ☐ over 50

*Years of teaching experience:

☐ less than 5 years

☐ 6-10 years

☐ 11-15 years

☐ 16-20 years

☐ more than 20 years

*Educational qualifications:

☐ BA in English language and literature

☐ BA in English Language/Field Teacher

☐ Higher Diploma

☐ MA/ MEd

☐ PhD/ EdD

☐ Other____

*Have you attended a course or a workshop on assessment and evaluation? *If yes, please specify the type of course (undergraduate/graduate/special) or the workshop (who provided it).*

*How do you perceive your knowledge about language assessment and evaluation?

☐ Excellent

☐ Very good

☐ Good

☐ Poor knowledge

☐ No knowledge

2: Purposes of Assessment

I use assessment in my classroom for the following purposes/reasons:					
	Always 100%	Usually 75%	Often 50%	Rarely 25%	Never 0%
1. To group my students for instruction purposes in my class.					
2. To obtain information on my progress.					
3. To plan my instruction.					
4. To diagnose strengths and weaknesses in my own teaching and instruction.					
5. To provide feedback to my students as they progress through the course.					
6. To motivate my students to learn.					
7. To make my students work harder.					
8. To diagnose strengths and weaknesses in my students.					
9. To formally document growth in learning.					
10. To determine the final grades for my students.					
11. To provide information to the school administration.					

Part 3: Methods of Assessment

A. Reading Assessment

I use the following methods to assess reading:

	Always	Usually	Often	Rarely	Never
	100%	75%	50%	25%	0%
Instructor-made					
Multiple-choice					
True/false					
Matching items					
Short answer items					
Sentence completion					
Cloze items					
Student-Conducted					
Student summaries of what is read					
Student portfolio					
Peer assessment					
Self-assessment					
Standardized Tests					

B. Writing Assessment

I use the following methods to assess writing:

	Always	Usually	Often	Rarely	Never
	100%	75%	50%	25%	0%
Instructor-Made					
Short essay					
Editing a sentence/ a Paragraph					
Long essay					
Multiple-choice items to identify grammatical error(s) in a sentence					
Matching items					
True/false					
Student-Conducted					
Self-Assessment					
Peer Assessment					
Student Portfolio					
Student journal					
Standardized Tests					

C. Listening and Speaking

I use the following methods to assess speaking and listening:

Always	Usually	Often	Rarely	Never
100%	75%	50%	25%	0%

Instructor-Made

Multiple-choice test items following a
listening passage

Take notes

Prepare summaries of what is heard

Student-Conducted

Oral interviews= Dialogues

Oral discussion with each student

Provide an oral description of an
event or object

Oral presentations

Follow directions given orally

Give oral directions

Retell a story after listening to a
passage

Oral reading/dictation

Self-assessment

Peer assessment

Standardized Listening Test

Standardized Speaking Test

PART 4: Beliefs about Language Assessment

	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
1- Assessment should be always associated with marks.					
2- Assessment results should be used to modify instruction to better support students' learning					
3-Students should be given opportunities to reflect on their learning through peer feedback and self-assessment.					
4- Students should be given feedback which enables them to know how to improve their work and take their learning forward.					
5- Teachers should negotiate the assessment criteria with their students in advance.					
6- Teachers should first identify the purpose of the assessment when they design the assessment task/test.					
7- Students will learn the most and perform their best if they know in <i>advance</i> the criteria that will be applied in judging a performance.					
8- Involving students in self- and peer assessment activities can enhance their achievement and motivation.					

* When you give feedback to your students during the course, how do you provide that feedback? Please check all that apply.

- a. Verbal feedback
- b. Checklist
- c. Written comments
- d. Conference with student
- e. Total test score
- f. A letter grade
- g. Other _____





Appendix B: Interview Questions

- 1- In your opinion, what is the purpose of classroom assessment?
- 2- In your opinion, what are the factors that influence classroom assessment positively or negatively?
- 3- What types of assessment methods do you use to assess students' reading competence in English language in your classroom? What is the most frequent one you use and why?
- 4- What types of assessment methods do you use to assess students' writing competence in English language in your classroom? What is the most frequent one you use and why?
- 5- What types of assessment methods do you use to assess students' speaking competence in English language in your classroom? What is the most frequent one you use and why?
- 6- What types of assessment methods do you use to assess students' listening competence in English language in your classroom? What is the most frequent one you use and why?
- 7- Do you use self-assessment, peer assessment and portfolios in the classroom? Why? Why not?
- 8- Do you inform the students about the assessment criteria before the task? If so, how?
- 9- Do you provide students with feedback in the class /outside of the class? What methods do you use for that?
- 10- For which purposes do you use the assessment results?



Appendix C: Ethics Committee Decision

Evrak Tarih ve Sayısı: 21.05.2024-120043



T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ REKTÖRLÜĞÜ
Lisansüstü Eğitim Enstitüsü Müdürlüğü



Sayı :E-88083623-020-120043
Konu :Etik Onayı Hk.

21.05.2024

Sayın Merve SARGIN

Tez çalışmanızda kullanmak üzere yapmayı talep ettiğiniz anketiniz İstanbul Aydın Üniversitesi Eğitim Bilimleri Etik Kurul Komisyonu'nun 15.05.2024 tarihli ve 2024/04 sayılı kararıyla uygun bulunmuştur.

Bilgilerinize rica ederim.

Dr. Öğr. Üyesi Nuri Gökmen KARAKIRAZ
Müdür Yardımcısı

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu : BSM4VT0PAP Pın Kodu : 09482

Belge Takip Adresi : <https://www.turkiye.gov.tr/istanbul-aydin-universitesi-ebys?>

Adres : Beşyol Mah. İnönü Cad. No:38 Sefaköy , 34295 Küçükçekmece / İSTANBUL

Telefon : 444 1 428

Web : <http://www.aydin.edu.tr/>

Kep Adresi : iau.yazisleri@iau.hs03.kep.tr

Bilgi için : Tuğba SUNNETCI

Unvanı : Yazı İşleri Uzmanı

Tel No : 31002



RESUME

Personal Information:

Name Surname: Merve SARGIN

Education:

MA: 2021-2024- Istanbul Aydin University: Institute of Graduate Studies English Language Department (with Thesis)

MA: 2011- 2017- Istanbul University - Institute of Social Sciences: Woman's Studies (with Thesis)

BA: 2005-2010- Istanbul University - Department of English Language and Literature

Work Experience:

2019- present: Yasar University Preparatory English Program, Bornova, Istanbul

2019- 2021: Dogus University Preparatory English Program, Umraniye, Istanbul

2013-2019: Kavram Vocational High School, Cengelkoy, Istanbul

Languages:

- TURKISH: Mother Tongue.

-ENGLISH: Fluent

-GERMAN: Beginner

Nationality:

Turkish.

