



**T.C.
PAMUKKALE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANABİLİM DALI
İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI
DOKTORA TEZİ**

**THE EFFECT OF ALTERNATIVE ASSESSMENT
APPLICATIONS ON STUDENTS' ACADEMIC
ACHIEVEMENT AND ATTITUDES TOWARDS LEARNING
ENGLISH IN ENGLISH PREPARATORY SCHOOL**

Hüsnü GÜMÜŞ

Denizli – 2024

T.R.
PAMUKKALE UNIVERSITY
THE INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM
DOCTORAL DISSERTATION

THE EFFECT OF ALTERNATIVE ASSESSMENT APPLICATIONS
ON STUDENTS' ACADEMIC ACHIEVEMENT AND ATTITUDES
TOWARDS LEARNING ENGLISH IN ENGLISH PREPARATORY
SCHOOL

Hüsnü GÜMÜŞ

Supervisor

Prof. Dr. Turan PAKER

ETİK BEYANNAMESİ

Pamukkale Üniversitesi Eğitim Bilimleri Enstitüsü'nün yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında; tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi; görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu; başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu; atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi; kullanılan verilerde herhangi bir tahrifat yapmadığımı; bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı beyan ederim.

Hüsnü GÜMÜŞ

ACKNOWLEDGEMENTS

First of all, I would like to express my heartfelt appreciation and gratitude to my supervisor Prof. Dr. Turan Paker for his endless patience, invaluable guidance, and support throughout my study. He was always accessible and eager to share his professional expertise with me when I needed help with my dissertation. It was a great pleasure to work with him.

In addition, I am also deeply indebted to Prof. Dr. Recep Şahin Arslan and Assoc. Prof. Dr. Perihan Korkut for the invaluable knowledge, guidance, and motivational attitude that they have provided to me during the Thesis Supervision Committee meetings . It was an honor for me to be guided by such excellent academicians. I also owe many thanks to the jury members Prof. Dr. Gonca Yangın Ekşi and Assoc. Prof. Dr. Çağla Atmaca for their invaluable contributions to my dissertation. I would also like to thank Prof. Dr. Demet Yaylı and Assist. Prof. Dr. Filiz Rızaoğlu, who made great contributions to me during my doctoral courses.

I would like to extend my endless thanks to my dear students who had a very busy assignment period during the alternative assessment implementation period. Without their voluntary efforts, I would not have been able to collect the data needed for this study.

I am also grateful to Şeyma Aktaş, a decent friend, a hardworking teacher and researcher and above all a wonderful person, for her support and encouragement throughout the whole process of writing my PhD thesis. I would also like to thank Assoc. Prof. Dr. İbrahim Halil Yurdakal for his help when I needed support in the analysis of quantitative data. I also owe many thanks to my colleagues who facilitated the data collection process with their collaboration and support.

Last but not least, I am grateful to my dear family for their support throughout my education. I have always felt lucky to have such a great family. Finally, my special thanks go to my beloved wife, Sevde Nur Gümüş, for her continuous support, patience, and encouragement throughout this long and difficult process. She always believed that I would succeed in completing this dissertation. Thank you for being there for me.

ÖZET

Alternatif Değerlendirme Uygulamalarının, İngilizce Hazırlık Okulu Öğrencilerinin Akademik Başarısı ve İngilizce Öğrenmeye yönelik Tutumları Üzerindeki Etkisi

GÜMÜŞ, Hüsnü

Doktora Tezi, Yabancı Diller Eğitimi Anabilim Dalı
İngiliz Dili Eğitimi Bilim Dalı
Tez Danışmanı: Prof. Dr. Turan PAKER
Mayıs 2024, 188 Sayfa

Bu karma yöntemli çalışmanın amacı, alternatif değerlendirme yöntemlerinin uygulanmasının bir grup İngilizce'yi yabancı dil olarak öğrenen Türk öğrencilerin akademik başarısı üzerindeki etkisini bulmak ve bu öğrencilerin alternatif değerlendirme yöntemlerinin uygulanmasından önce ve sonra İngilizce öğrenmeye yönelik tutumlarını araştırmaktır. Araştırmanın katılımcıları kolaylıkla bulunabilen örnekleme yöntemiyle seçilmiş olup çalışmada 75 hazırlık sınıfı öğrencisi bulunmaktadır. Deney grubunda 38 öğrenci yer alırken, kontrol grubunda ise 37 katılımcı yer almıştır. Deney grubundaki katılımcılar, süreç boyunca dört becerinin (yani okuma, dinleme, konuşma ve yazma) entegre edildiği çeşitli alternatif değerlendirme teknikleri ve ödevlere dahil olmuş ve performansları önceden belirlenmiş değerlendirme listeleri ve kontrol listeleri aracılığıyla sürece dayalı bir şekilde değerlendirilmiştir. Kontrol grubundaki katılımcılar ise süreç içerisinde geleneksel bir öğretim deneyimi yaşamıştır. Her beceri dersi için ders kitaplarındaki alıştırmaları tamamlamışlar ve ödev olarak verilen çalışma sayfalarını yapmışlardır. Performansları standartlaştırılmış yazılı sınavlar, testler ve kısa sınavlar gibi geleneksel değerlendirme yöntemleriyle değerlendirilmiştir. Nicel veriler tutum ölçeği, geleneksel sınavlar, alternatif değerlendirme araçları yoluyla toplanırken, nitel veri toplama araçları yarı yapılandırılmış görüşmeler ve öğrenme günlüklerini içermektedir. Ölçekten, geleneksel sınavlardan ve alternatif değerlendirme araçlarından elde edilen nicel veriler SPSS 20 kullanılarak analiz edilmiştir. Nitel verilerin analizi ise içerik analizi aşamaları takip edilerek gerçekleştirilmiştir. Araştırmanın nicel bulguları, bazı alternatif değerlendirme araçlarının uygulanmasının deney grubundaki katılımcıların akademik başarıları üzerinde olumlu etki yarattığını ortaya koymuştur. Benzer şekilde nitel bulgular, deney grubundaki katılımcıların uygulama süreci sonrasında hem alternatif değerlendirme yöntemlerine hem de İngilizce öğrenmeye yönelik olumlu tutumlar sergilediklerini göstermiştir.

Anahtar Kelimeler: Ölçme ve değerlendirme, alternatif ölçme araçları, İngilizce dil öğretimi, İngilizce öğrenmeye yönelik tutum, hazırlık sınıfı öğrencileri

ABSTRACT

The Effect of Alternative Assessment Applications on Students' Academic Achievement and Attitudes towards Learning English in English Preparatory School

GÜMÜŞ, Hüsnü

Ph.D. Dissertation in Department of Foreign Languages
English Language Teaching
Supervisor: Prof. Dr. Turan PAKER
May 2024, 188 Pages

The purpose of this mixed-method study is to find out the effect of alternative assessment methods on a group of Turkish EFL students' academic achievement and investigate the attitudes of these students towards learning English before and after the implementation of alternative assessment methods. The participants of the study were selected through the convenience sampling method and included 75 preparatory class students. There were 38 participants in the experimental group, while the control group consisted of 37 participants. The participants in the experimental group were engaged in several alternative assessment techniques and tasks where four skills (i.e., reading, listening, speaking, and writing) were integrated and their performances were assessed in a formative manner through predetermined rubrics and checklists throughout the process. On the other hand, the participants in the control group experienced conventional instruction during the process. For each skill lesson, they completed the existing exercises in their textbooks and did the worksheets assigned as homework. Their performances were assessed through traditional assessment methods such as standardized written exams, tests, and quizzes. The quantitative data were collected through an attitude scale, traditional exams, and alternative assessment tools while the qualitative data collection tools included semi-structured interviews and learning journals. The quantitative data gathered from the scale, traditional exams and alternative assessment tasks were analyzed using SPSS 20. On the other hand, the analysis of qualitative data was performed following the stages of content analysis. The quantitative findings of the study revealed that applying the alternative assessment tools had a positive impact on the academic achievement of the participants in the experimental group. Likewise, the qualitative findings indicated that the participants in the experimental group demonstrated favorable attitudes towards both alternative assessment methods and learning English after the treatment process.

Keywords: Testing and assessment, alternative assessment tools, English language teaching, attitude towards learning English, preparatory school students

TABLE OF CONTENTS

ETİK BEYANNAMESİ	iv
ACKNOWLEDGEMENTS	v
ÖZET	vi
ABSTRACT.....	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiv
CHAPTER 1: INTRODUCTION	1
1.1. Background to the Study	1
1.2. Statement of the Problem	4
1.3. Purpose of the Study	8
1.4. Significance of the Study	9
1.5. The Research Questions	10
1.6. Limitations of the Study	12
1.7. Assumptions of the Study	12
CHAPTER 2: LITERATURE REVIEW	13
2.1. Communicative Language Teaching.....	13
2.2. Constructivism	14
2.3. Bloom’s Taxonomy.....	16
2.4. Assessment.....	17
2.4.1. Assessment in English Language Teaching	18
2.5. Assessment Types	19
2.5.1. Traditional Assessment	20
2.5.1.1. Criticism of traditional assessment.	21
2.5.2. Alternative Assessment	23
2.5.2.1. Benefits of alternative assessment.	23
2.5.2.2. Challenges of alternative assessment.....	25
2.5.2.3. Solutions for the challenges.....	27
2.6. Types of Alternative Assessment.....	29
2.6.1. Portfolios	30

2.6.2. Performance Assessment.....	32
2.6.3. Learning/Reflective Journals.....	33
2.6.4. Conferences.....	34
2.6.5. Self-Assessment.....	34
2.6.6. Peer Assessment.....	36
2.7. Tools for Alternative Assessment.....	37
2.7.1. Grading Rubrics and Checklists.....	37
2.8. Previous Studies on Alternative Assessment.....	38
2.8.1. Studies Conducted on Students' Academic Achievement.....	38
2.8.1.1. Portfolio studies.....	38
2.8.1.2. Self-Assessment and peer assessment studies.....	44
2.8.2. Studies Conducted on Students' Perceptions of Alternative Assessment.....	49
CHAPTER 3: METHODOLOGY.....	59
3.1. Research Design.....	59
3.2. Setting and Participants.....	61
3.3. Data Collection.....	63
3.3.1. Data Collection Instruments.....	63
3.3.1.2. Traditional exams.....	64
3.3.1.3. Alternative assessment tasks.....	65
3.3.1.4. Semi-Structured interviews.....	67
3.3.1.5. Learning /reflective journals.....	68
3.4. Data Collection Process.....	69
3.4.1. The Pilot Study.....	69
3.4.2. The Main Study.....	71
3.4.2.1. Treatment process for the experimental group.....	72
3.4.2.2. The process for the control group.....	78
3.5. Data Analysis.....	80
CHAPTER 4: FINDINGS.....	82
4.1. RQ1. Is There a Statistically Significant Difference between the Overall Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group?.....	83
4.1.1. RQ1a. Is There a Statistically Significant Difference between the Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Writing Skills?.....	83

4.1.2. RQ1b. Is There a Statistically Significant Difference between the Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Reading Skills?	84
4.1.3. RQ1c. Is There a Statistically Significant Difference between the Alternative Assessment scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Listening Skills?	84
4.1.4. RQ1d. Is There a Statistically Significant Difference between the Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Speaking Skills?	85
4.2. RQ2. Is There a Statistically Significant Difference between the Overall Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group?.....	86
4.2.1. RQ2a. Is There a Statistically Significant Difference between the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Writing Skills?	86
4.2.2. RQ2b. Is There a Significant Difference between the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Reading Skills?	87
4.2.3. RQ2c. Is There a Statistically Significant Difference between the Traditional Assessment scores of the Participants in the Experimental Group and the Control Group in terms of Listening Skills?	87
4.2.4. RQ2d. Is There a Statistically Significant Difference Between The Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in Terms of Speaking Skills?	88
4.3. RQ3. What are the Participants' Attitudes towards Learning English Before and After the Implementation of Alternative Assessment Applications?.....	88
4.3.1. RQ3a. Is There a Statistically Significant Difference between the Pretest and Posttest Scores of the Participants in the Control Group in terms of Their Attitudes towards Learning English?	90
4.3.2. RQ3b. Is There a Statistically Significant Difference between the Pretest and Posttest Scores of the Participants in the Experimental Group in terms of Their Attitudes towards Learning English ?	92
4.4. Findings from the Interviews	94
4.4.1. Effective Learning Process.....	96

4.4.2. Fairer Assessment.....	98
4.4.3. Affective Contribution.....	99
4.4.4. Suggestions for Future.....	101
4.4.5. Comparison of Two Methods.....	102
CHAPTER 5: DISCUSSION, CONCLUSION, AND SUGGESTIONS.....	104
5.1. Discussion	104
5.1.1. Discussion of the First Research Question and Its Sub-problems:.....	104
5.1.2. Discussion of the Second Research Question and Its Sub-problems:	109
5.1.3. Discussion of the Third Research Question and Its Sub-problems:	117
5.2. Conclusion, Suggestions and Pedagogical Implications.....	120
5.2.1. Pedagogical Implications.....	124
5.2.1.1. Suggested guideline for alternative assessment implementation.....	125
5.2.2. Suggestions for Further Research.....	128
REFERENCES	130
APPENDICES	149
Appendix A	149
Appendix B	155
Appendix C	157
Appendix D	161
Appendix E.....	164
Appendix F.....	165
Appendix G.....	166
Appendix H	172

LIST OF TABLES

Table 2.1. <i>Paradigm Shift from Traditional Assessment to Alternative Assessment</i>	22
Table 2.2. <i>Traditional and Alternative Assessment</i>	25
Table 3.1. <i>Traditional Assessment Tools and Their Percentages</i>	65
Table 3.2. <i>Schedule of the Assigned Tasks for the Experimental Group (Module B1)</i>	76
Table 3.3. <i>Schedule of the Assigned Tasks for the Experimental Group (Module B1+)</i> ...	77
Table 3.4. <i>Summary of the Research Design, Participants and Procedures</i>	81
Table 4.1. <i>Normality Test Results of the Participants</i>	82
Table 4.2. <i>Comparison of the Participants' Placement Test Scores.</i>	82
Table 4.3. <i>Comparison of the Overall Alternative and Traditional Assessment Scores of the Participants in the Experimental Group</i>	83
Table 4.4. <i>Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Writing Skills</i>	83
Table 4.5. <i>Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Reading Skills</i>	84
Table 4.6. <i>Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Listening Skills</i>	85
Table 4.7. <i>Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Speaking Skills</i>	85
Table 4.8. <i>Comparison of the Overall Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group</i>	86
Table 4.9. <i>Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Writing Skills</i>	86
Table 4.10. <i>Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Reading Skills</i>	87
Table 4.11. <i>Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Listening Skills</i>	87
Table 4.12. <i>Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Speaking Skills</i>	88
Table 4.13. <i>Normality Test Results of the Attitude Scale</i>	89

Table 4.14. <i>Comparison of the Pretest and Posttest Scores of the Participants in the Control Group on the Attitude Scale</i>	90
Table 4.15. <i>Descriptive Statistics of the Participants in the Control Group on the Attitude Scale (Item-based)</i>	91
Table 4.16. <i>Comparison of the Pretest and Posttest Scores of the Participants in the Experimental Group on the Attitude Scale</i>	93
Table 4.17. <i>Descriptive Statistics of the Participants in the Experimental Group on the Attitude Scale (Item-based)</i>	93
Table 4.18. <i>Themes and Categories Regarding Alternative Assessment Applications</i>	96
Table 5.1. <i>Suggested Alternative Assessment Implementation Guideline</i>	128

LIST OF FIGURES

<i>Figure 2.1.</i> Principles of constructivism	15
<i>Figure 2.2.</i> Bloom's taxonomy.....	16
<i>Figure 2.3.</i> Alternative assessment tools and methods.....	30
<i>Figure 3.1.</i> A Sequential explanatory design	61
<i>Figure.4.1.</i> Histogram for the normality of the attitude scale	90

CHAPTER 1: INTRODUCTION

This chapter presents the background to the study and introduces the statement of the problem. In addition, the purpose of the study, research questions, significance and limitations of the study are presented.

1.1. Background to the Study

Assessment and evaluation are concepts that have always been a focus of research studies in the English Language Teaching (ELT) field since they are viewed as indispensable components of language teaching programs and have the potential to shape and influence the instruction and assessment policies dramatically (Dochy, 2001; Remesal, 2011). In fact, assessment plays a crucial role in language teaching, because it informs teachers and other stake- holders such as school administrators and parents about the extent to which learners have achieved what they have been taught in class.

To emphasize the direct and significant relationship between teaching and assessment, DiRanna et al. (2008) state that “assessment and instruction are two sides of the same coin” (p.22). Regarding the role of assessment in language teaching, Bachman (1990) also points out that language tests, regardless of whether they are formal or informal, provide teachers with information related to the strengths and weaknesses of their teaching practices, allow them to make the necessary adjustments to their ways of teaching when necessary, and monitor their students’ progress and general improvement in class.

Due to the importance of testing and assessment, people who are responsible for designing curriculum, syllabi, and course materials need to consider the existing assessment tools and strategies in their institutions so that learners can be assessed through appropriate assessment methods and have a high-quality education (Popham, 2009). Due to the strong relationship between testing and teaching, stakeholders need to make sure that learners are assessed through appropriate testing tools to accomplish the goals of their curriculums. The importance of the congruence between testing tools and language content has been emphasized through the term washback. Washback is defined as the impact of assessment procedures on teaching and learning practices (Hughes, 2003). According to Taylor (2005), when assessment procedures match a curriculum’s objectives or goals, positive or beneficial washback occurs. For instance, if students must take a speaking test at the end of the academic year, the teacher will focus on improving students’ speaking skills throughout the

term and design the activities and materials accordingly. Contrary to positive washback, negative washback happens when the assessment procedures do not serve a curriculum's objectives or goals. For example, if an institution sets a series of objectives to improve the communicative skills of learners but assesses the learners through multiple-choice or true-false test items, this creates a negative washback effect since the teacher will focus on improving learners' test performances rather than their speaking skills (Hughes, 2003).

In the last decade of the 20th century, the prevalent paradigm which emphasized knowledge transfer was replaced by new paradigms such as learner-centered applications and educational practices which aimed to develop learners' competences (Hamayan, 1995). Following the recent educational reform movements, teaching practices in language settings have shifted their aims from merely transferring knowledge to learners to what skills and competencies they need to possess to be successful in the real world. In other words, learners are actively engaged in the learning process in a social environment and take responsibility for their own learning. Accordingly, professionals who work in the language teaching field aim to equip learners with critical thinking skills and try to enable them to solve problems through collaborating and communicating with others efficiently (Woolfolk, 2005).

Due to the paradigm shift that resulted in learner-centered approaches in ELT, assessment applications have also been profoundly influenced (Jacobs & Farrell, 2001). It has been claimed that traditional assessment applications and devices fail to assess the higher-level thinking skills of learners since they are unable to reflect the underlying knowledge of learners (Lizasoain & Zárate, 2014). Therefore, language teaching professionals have adopted new approaches and developed more appropriate tools of student assessment to demonstrate what students are learning and what they are able to do with their underlying knowledge in real-life situations (Jacobs & Farrell, 2001). These new assessment methods, commonly defined as alternative or authentic measuring tools, require learners to demonstrate higher-order thinking skills in certain forms such as presenting, reporting, demonstrating, constructing, explaining, or completing a task or project (Dikli, 2003; Koh, 2017). The accomplishment of such objectives obviously requires an integrated approach where language skills will be acquired and developed during the learning process and assessed through interrelated tasks (Frey & Schmitt, 2007).

It is a well-known fact that an educational institution's principal goal is to provide a quality education for its learners and prepare them for their professions, careers, and social life. Considering this goal, these institutions are obliged to adopt appropriate assessment applications to demonstrate what their students know and to what extent they are able to

apply the knowledge they have acquired during their classes. Regarding student achievement, tests have always been used as a tool to estimate learners' competence (Hancock, 1994). In today's competitive world, important life decisions are made based on test scores, and individuals are judged based on their performances in certain standardized, traditional tests. However, Tsagari (2004) argues that such tests could narrow the curriculum, restrict the use of different teaching methods, and even alter the instructional materials used in a specific institution. As a result, they could have a negative washback effect on learning (Hughes, 2003). According to Dikli (2003), such traditional assessment applications are unable to serve the real-life objectives of language learning since they are indirect and inauthentic. In addition, they only reflect what learners can achieve at a particular time rather than reflecting their real competence, and the score references obtained from them may not always be valid (Gottlieb, 2006).

Since a great deal of dissatisfaction has been observed related to the traditional assessment methods, people working in the education field have started to search for more learner-focused assessment methods (Gipps, 1994). As opposed to the use of traditional assessment, the scholars who were interested in assessment applications proposed a term called alternative assessment. According to Hancock (1994), alternative assessment can be described as a continuous process that engages both the student and teacher in the learning process and requires mutual negotiation in making decisions regarding the learner's progress. Such alternative assessment applications are utilized by language teachers since they assess not only the process but also the product of learning and provide continuous feedback for both the teachers and the students (Huang, 2016). Furthermore, they enable students to discover their own learning preferences, make them aware of their strengths and weaknesses, and encourage them to take responsibility for their own learning. (Crick & Yu, 2008). Regarding alternative assessment, McNamara (2001) states that it is a kind of assessment that differs from traditional paper and pencil tests since it requires learners to perform tasks by demonstrating skills and applying knowledge that cannot be assessed through multiple-choice tests. Therefore, it promotes and enhances meaningful learning and has a positive washback effect on education (Hamp-Lyons, 1997). Finally, alternative assessment methods are usually applied in a formative manner and enable learners to track and monitor their own progress in addition to assisting teachers to modify their teaching methods or materials if necessary.

In the Turkish context, foreign language education experienced a transformation in the light of the communication-oriented language curriculum reform that took place at the

end of the 20th century. Accordingly, foreign language assessment applications were also altered, and authentic and performance-based assessment methods started to be used by teachers (Kırkgöz, 2007). However, standardized high-stakes exams have always been dominant at every stage of education. Likewise, when the foreign language education context in preparatory schools is considered, common practices indicate that the process of assessment and evaluation has been dominated by traditional assessment tools in Türkiye. Institutions tend to apply traditional assessment tools such as pen and paper tests and exams that include multiple-choice, matching, true –false or gap-filling test items to measure their students' performance (Kırkgöz & Ağçam, 2012). However, they also indicate that these traditional assessment tools are unable to reflect the real-life language skills of students. Therefore, new or alternative forms of assessment which focus more on measuring learner's ability to use the language communicatively and holistically in real-life situations need to be used more effectively.

There have been many research studies regarding the role of alternative assessment in foreign language education (Burnaz, 2011; Cirit, 2014; Pesen, 2016; Özuslu 2018). Since the context of preparatory schools is unique, it is necessary to find out how alternative assessment tools could be utilized effectively to boost student achievement in these schools. In addition, it is widely accepted that students' attitudes towards learning English need to be improved in Türkiye (Altınar, 2018). Therefore, it is worth investigating whether students' attitudes towards learning could be improved using alternative assessment tools. In essence, this study aims at finding out the effect of applying alternative assessment tools on a group of Turkish EFL students' academic achievement and investigating the attitudes of these students towards learning English before and after the implementation of alternative assessment tools in a preparatory school in Türkiye.

1.2. Statement of the Problem

In today's globalizing world, it is of great importance for people to learn English and use it effectively in almost every part of life. Due to globalization, it has become a priority for many countries around the world to provide a quality English education for their citizens since it is a crucial tool to keep up with the developments in technology, education, economy, entertainment, and so on (Tsui & Tollefson, 2007). Therefore, educational policy makers have attached great importance to English education and searched for appropriate teaching methods and assessment procedures to meet the demands of the increasingly evolving world.

Radical changes have been observed in education due to the theory of constructivism, which has provided a general framework both for teaching and assessment practices all around the world. As for teaching, the constructivism theory indicates that a teaching method ought to be learner-centered rather than teacher-centered since learning is a social, interactive, and developmental process in which learners undertake an active role in constructing the meaning (Poehner, 2005). When it comes to assessment, the constructivism theory called for a shift from traditional standardized assessment to process-based performance assessment, which aims to provide evidence of learner progress on a regular basis (Stiggins, 2005).

As mentioned above, teacher-centered approaches started to be replaced by more student-centered ones, which has also altered foreign language assessment practices in a similar pattern (Grabin, 2007). Due to the inadequacies of traditional assessment techniques, educators started to develop new or alternative assessment techniques. While traditional assessment employs conventional or standardized methods of testing such as multiple choice, true/false, and gap filling, alternative assessment reflects the traits of communicative language teaching and includes process-based performance tasks. In other words, learners are engaged in meaningful real-life tasks in a formative manner. Therefore, it could be stated that using alternative assessment fosters learners' communicative competences, and thus has a positive washback effect in the EFL classroom (Duff, 2014).

As for the Turkish context, in the light of the above-mentioned approaches, foreign language education has undergone remarkable changes since the Education Reform that occurred in 1997. The most striking innovation which resulted from this reform is that foreign language education ought to be more communication oriented and learner centered. The 1997 foreign language curriculum also maintains that the real objective of foreign language education should be to improve the communicative skills of learners by integrating four language skills (i.e., listening, reading, speaking, and writing) so that language learners could achieve successful communication in the target language (MONE, 2001). It could be stated that the 1997 foreign language curriculum was a turning point in foreign language education since it paved the way for a more communicative foreign language teaching in Türkiye (Kırkgöz, 2005).

In line with the changes in foreign language education mentioned, there has also been a shift from traditional assessment practices to a more holistic and performance-based assessment in foreign language classrooms (Kırkgöz, 2007). This shift was in parallel with the changes that occurred in foreign language assessment practices all around the world.

However, when it comes to the implementation of these changes, it can be clearly seen that traditional one-shot language tests are still widely used in most of the schools in Türkiye though the curriculum states that these standardized assessment tools do not serve the main objective of English education (Paker, 2006). People who are in charge of assessment policies state that traditional assessment tools are practical, reliable, and easy to use compared to alternative assessment methods. However, these static assessment practices neglect the communicative aspect of language teaching and do not reflect the underlying knowledge of learners. According to Braun and Mislevy (2005), these standardized tests only measure learners' ability to recall information and lead to rote learning, and thus cannot promote or measure higher order thinking skills. Therefore, they might produce inaccurate results regarding the performance of learners (Barootchi & Keshavarz, 2002). In addition, they create a negative washback in teaching since teachers only focus on preparing their students for exams at the expense of students' communicative skills. In other words, these standardized tests restrict the curriculum and teachers in several ways (Hughes, 1990). In this sense, it could be stated that these traditional tests play a vital role in shaping teacher's practices in classrooms (Bachman & Palmer, 1996).

Despite the prevalence of traditional assessment applications, there has been a growing interest in the implementation of alternative assessment applications among language teachers in Türkiye. Alternative assessment applications encompass a wide range of tools such as checklists, journals, rubrics, self-reflection, teacher observations, learner portfolios, conferences, self-assessments, and peer assessments, and so on (Brown & Hudson, 1998). These new forms of assessment tools are needed since traditional assessment tools fail to reveal what learners can achieve in real life using their underlying language knowledge. Many language teachers in Türkiye complain that the existing high-stakes exams force students to memorize the rules of the language to pass some tests rather than using it for communicative purposes (Paker, 2006). To fill this gap, alternative assessment applications have come into play since they require students to think critically and accomplish certain tasks through the integration of skills. Another major problem of language assessment practices in Türkiye is that they are conducted in a summative manner. Therefore, teachers cannot determine the strengths or weaknesses of their students before it is too late. However, alternative assessment is conducted in a formative manner and both processes and products are considered while judging the performance of learners (Alderson & Banerjee, 2001; Guskey, 2003). In addition, alternative assessment applications cater for the individual needs and learning styles of students. Regarding this issue, Yıldırım and

Orsdemir (2013) state that alternative assessment promotes learning and increases the opportunity, equality, and autonomy in education for all students. Therefore, it is essential that English teachers in Türkiye utilize appropriate alternative assessment methods considering the objectives of their lessons.

All school of foreign languages in Türkiye have their own syllabi, teaching systems and assessment policies. Their main objective is to enable their students to accomplish the learning outcomes of the Common European Framework of Reference (henceforth CEFR). In CEFR, there are descriptors which demonstrate the expected outcomes for each level of proficiency (CoE, 2001). Since it is an internationally recognized reference, CEFR provides a standardization of language proficiency criteria and defines learners' language levels across different institutions in Europe (Mirici & Kavaklı, 2017). Therefore, schools of foreign languages in Türkiye have redesigned their curriculum, teaching methods, and assessment policies based on CEFR (Yüce & Mirici, 2019).

In the School of Foreign Languages at Pamukkale University, students come from many different degree programs including both 100% and 30% English Medium Instruction (EMI) students as well as the ones who take preparatory school education electively. In the school, the foreign language learning program is divided into four modules which are identified as A1, A2, B1, and B1+. Each module lasts eight weeks with 24 hours of English lessons per week. In a weekly schedule, five hours are allocated to reading and writing courses each while speaking and listening courses are allocated three and two hours respectively. Lastly, nine hours are allocated to core lessons. In each module, students are required to take a midterm and a final exam in pen and paper form. These exams consist of four sections such as reading, writing, listening and language use. Along with the midterm and final exams, students also take a quiz for each skill during the module. They also take a speaking quiz in the middle of the module and another speaking exam at the end of the module. However, these speaking exams only last about five to ten minutes and include question items that were provided to the students beforehand. As a result, students tend to memorize the answers to these questions prior to the exam. Although this provides an opportunity for the students to be engaged in the target language, they usually end up with poor performance in the long term. To be considered successful and progress to the next level, a student needs to obtain a minimum of 70 points on average. While calculating this average score, 90% of standardized pen-and-paper exams (i.e., quizzes, midterm, and final exams) are taken into account while only 10% is dedicated to in-class performance assessment. Additionally, the in-class assessment of students and the tasks assigned vary

enormously depending on the language instructors. In other words, there is not a systematic or standardized in-class performance assessment of students based on predetermined scoring criteria.

As one can see, students have a tight schedule of exams during each module and both students and teachers aim to achieve high scores in these exams. As a result, teachers tend to ignore the communicative purpose of language teaching and struggle to prepare their students for these traditional exams. Likewise, students demand exam-oriented instruction from their teachers, which leads to a negative washback effect on language teaching. Apart from these, some students cannot perform well in these standardized exams due to outside factors such as family problems, sickness, time constraints or other personal issues. Therefore, these standardized one-shot exams, might not give teachers an accurate performance of their students. Furthermore, due to the pressure of sit-down exams, students usually develop a negative attitude towards learning English, and some of them even drop out of school in the middle of the academic year. However, if their performances were assessed through alternative assessment tools in a formative manner, they could take the necessary precautions before it is too late. Moreover, due to the variety of tasks and activities in alternative assessment, students would be more motivated to learn English. Although traditional, standardized assessment methods are preferred for their validity, reliability, and practicality, these assessment applications view language learning from a technical perspective rather than a functional one. As a language instructor working at a school of foreign languages, I believe that every language curriculum needs to include a variety of assessment methods rather than depending on only standardized sit-down exams. Therefore, within the scope of this dissertation, alternative assessment applications are suggested as an additional medium to foster language learning with a communicative and functional purpose through the integration of four skills. It is also expected that employing such methods will improve the attitudes of learners towards learning English at the school of foreign languages.

1.3. Purpose of the Study

In a general sense, in this dissertation, we aim to present a rationale and describe an implementation process for the systematic use of alternative assessment applications along with standardized exams in EFL classrooms. As a result of the paradigm shift that resulted in learner-centered approaches in ELT, there has been a great deal of dissatisfaction with standardized, one-shot assessment methods, which merely assess to what extent learners can

memorize or recall the information they have learned during their lessons (Çakır, 2013). Therefore, it has been widely acknowledged that new or alternative assessment methods are needed to assess students' language skills through meaningful tasks and gather evidence regarding their progress in language learning (Brown & Hudson, 1998; Stiggins, 2005; Sidek, 2012). Even though a more constructivist approach has been adopted in foreign language education in Türkiye, classroom practices and assessment procedures are still traditional in Türkiye (Kırkgöz, 2007). The content of such standardized exams promotes rote learning and does not match the learning outcomes of CEFR, which views language learning and assessment as a process rather than one-shot practice. In addition, traditional assessment methods have some negative features such as time pressure, exam anxiety, limited content, and so on. Therefore, students feel less motivated in class and might develop a negative attitude towards studying the subject. As Lizzio and Wilson (2013) point out, assessment methods have a vital role in the motivation level of learners and their engagement in the lesson. It is expected that applying certain alternative assessment methods, which assess learners' performances in a formative manner through meaningful tasks, could create a more positive learning atmosphere in classrooms. Therefore, this study aims at finding out the effect of applying alternative assessment tools on a group of Turkish EFL students' academic achievement and investigating the attitudes of these students towards learning English before and after the implementation of alternative assessment tools in a preparatory school in Türkiye.

1.4. Significance of the Study

Although there have been many attempts to improve the conditions in foreign language education, the expected outcomes are still far from being attained in schools in Türkiye (Kırkgöz, 2007; Paker, 2006). The current study assumes that traditional approaches in English language teaching and assessment systems have adverse effects on learners. Since knowing English is assumed to be equivalent to getting high test scores, both teachers and students spend a great deal of time and energy on passing tests and getting good grades.

It is widely acknowledged that tests have a huge impact on teaching practices, and what is asked in tests tends to shape teaching methods and materials. It is an undeniable fact that high-stakes language exams, whether they are proficiency or achievement tests, are still traditional, and thus, do not serve the communicative purposes of language teaching. In addition, these traditional tests put pressure on learners and discourage them from learning

the language for real-life purposes. Therefore, the current study aims to suggest some alternative assessment applications and enhance learners' language skills as well as improving their attitudes towards the language.

In the relevant literature, many studies have been conducted on the use of alternative assessment in EFL settings (Kızıl, 2019; Özer & Tanriseven, 2016; Özuslu, 2018; Pesen, 2016). Most of these studies attempted to investigate the effects of alternative assessments on only one specific language skill such as speaking, writing, listening, or reading. On the other hand, some other studies merely tried to explore the perceptions of teachers or students about alternative assessment tools. However, the current study differs from the previous studies in that it aimed to reveal the effects of alternative assessment on a group of EFL learners' overall academic achievement as well as their progress in four language skills. What is more, it aimed to reveal whether certain alternative assessment applications affect the attitudes of these learners towards learning English. Most importantly, the study aims to provide a basic alternative assessment implementation guideline for other schools of foreign languages in Türkiye so that they could reduce the ratio of traditional exams in their institutions by applying alternative assessment methods in a more systematic way.

1.5. The Research Questions

The present study aims at finding out the effect of applying alternative assessment tools on a group of EFL students' academic achievement and investigating the attitudes of these students towards learning English before and after the implementation of alternative assessment tools in a preparatory school. To achieve this purpose, the following research questions were posed:

- 1.** Is there a statistically significant difference between the overall alternative assessment scores and traditional assessment scores of the participants in the experimental group?
 - 1a.** Is there statistically a significant difference between the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of writing skills?
 - 1b.** Is there a statistically significant difference between the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of reading skills?

- 1c.** Is there a statistically significant difference between the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of listening skills?
- 1d.** Is there a statistically significant difference between the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of speaking skills?
- 2.** Is there a statistically significant difference between the overall traditional assessment scores of the participants in the experimental group and the control group?
- 2a.** Is there a statistically significant difference between the traditional assessment scores of the participants in the experimental group and the control group in terms of writing skills?
- 2b.** Is there a statistically significant difference between the traditional assessment scores of the participants in the experimental group and the control group in terms of reading skills?
- 2c.** Is there a statistically significant difference between the traditional assessment scores of the participants in the experimental group and the control group in terms of listening skills?
- 2d.** Is there a statistically significant difference between the traditional assessment scores of the participants in the experimental group and the control group in terms of speaking skills?
- 3.** What are the participants' attitudes towards learning English before and after the implementation of alternative assessment applications?
- 3a.** Is there a statistically significant difference between the pretest and posttest scores of the participants in the control group in terms of their attitudes towards learning English?
- 3b.** Is there a statistically significant difference between the pretest and posttest scores of the participants in the experimental group in terms of their attitudes towards learning English?

1.6. Limitations of the Study

The current study aims at finding out the effect of applying alternative assessment tools on a group of Turkish EFL students' academic achievement and investigating the attitudes of these students towards learning English before and after the implementation of alternative assessment tools in a preparatory school. However, the study has several limitations, which can pave the way for further studies in the field. To start with, the findings of the present study cannot be generalized to other schools of foreign languages in Türkiye since each institution has its own context. In addition, the number of participating students could be regarded as a limitation. A further study including more students might yield different results. Finally, the study was limited to only one instructor: the researcher of the study. A study including more instructors might produce different results.

1.7. Assumptions of the Study

For the present study, it is assumed that:

1. The sample selected for data collection represents the whole population.
2. The participants in both control and experimental group gave sincere answers to the questions on the attitude scale as they voluntarily participated in the study.
3. The participants in the experimental group took the process seriously and completed the alternative assessment tasks as they were instructed.
4. The participants in the experimental group gave sincere answers to the interview questions.

Considering these assumptions, the researcher conducted the research by completing each phase of the study carefully.

CHAPTER 2: LITERATURE REVIEW

This chapter presents the theoretical basis of alternative assessment and introduces basic definitions and concepts related to assessment. In addition, it provides a review of previously conducted studies on alternative assessment applications. Via incorporating the related studies into the theoretical framework, the significance of alternative assessment applications in EFL classrooms has been highlighted.

2.1. Communicative Language Teaching

Alternative assessment has its roots in Communicative Language Teaching (CLT henceforth) since it aims to improve the communicative competence of learners. CLT, which is viewed as an approach in language teaching, tries to compensate for the shortcomings of the traditional language teaching methods such as the Grammar Translation Method and Audio-Lingual Method. Its main aim is to improve learners' abilities to use English in real-life situations (Littlewood, 2007). Therefore, a shift has been observed in language classrooms from mechanical, drill-based activities to meaningful, communication-oriented activities so that students' language skills can improve naturally. In essence, CLT enables learners to develop their communicative competences in authentic and real-life contexts (Larsen-Freeman, 2003). Therefore, language curriculums, teaching materials, and assessment procedures have been redesigned in the light of CLT in many countries.

In CLT classrooms, teachers are expected to undertake several roles to facilitate learning. They are supposed to act as a resource, a counselor, a guide, a motivator, and a researcher. (Richards & Rodgers, 2001). They also monitor students' improvements throughout the learning process and provide constructive feedback when necessary (Dewey, 1916; Piaget, 1973; Vygotsky, 1978). On the other hand, in CLT classrooms, priority is given to learner engagement, and it is mostly the responsibility of learners to manage their own learning processes since they are aware of the objectives of the lessons and can address their own learning needs (Larsen-Freeman, 1986; Richards, 2001).

The main purpose of CLT is to equip learners with communicative competence. Basically, communicative competence consists of our knowledge of how to use the language properly depending on time, place, and to whom we are using it because there is an interdependence of language and communication to accomplish different acts (Fasold & Connor-Linton; Larsen-Freeman, 2000). According to Canale and Swain (1980),

communicative competence encompasses grammatical competence, discourse competence, social-cultural competence, and strategic competence. Grammatical competence refers to the recognition of the lexical, morphological, syntactic, and phonological features of a language, while discourse competence is concerned with the connectedness of a series of utterances in a written or oral text such as a poem or telephone conversation. On the other hand, social-cultural competence requires learners to understand the social context where language is used. It also considers the roles of the speakers, the information they share and the purpose of the interaction. Finally, strategic competence serves as a strategy that helps learners to cope with unexpected problems in an unfamiliar context. As one can see, a language learner needs to have communicative competence to use the language appropriately and effectively in different contexts. To achieve this, the teacher needs to adopt teaching techniques that are designed to encourage learners to be involved in the authentic, functional use of language and employ a variety of alternative assessment techniques to give learners enough opportunity to practice the language in pragmatic meaningful tasks (Brown, 2004).

2.2. Constructivism

Alternative assessment is also grounded in the framework of the constructivist theory. Constructivism refers to a learning theory which is based on a learner-centered teaching approach whereby learners undertake an active role in exploring and interpreting their learning experiences (Estrin, 1993; Shepard, 2000). Constructivist theory maintains that learners should not be viewed as passive recipients of knowledge. Instead, they should be actively involved in constructing their own knowledge, engaged in discourse, managing their learning experiences, and need to make connections between their already existing knowledge and new information (Driscoll, 2000; Kaufman, 2004; Shepard, 2000). The constructivist theory has some principles, and knowledge of these principles provides a guideline for teachers to create a constructivist learning atmosphere in their teaching environment. The principles are shown in Figure 1 as follows:

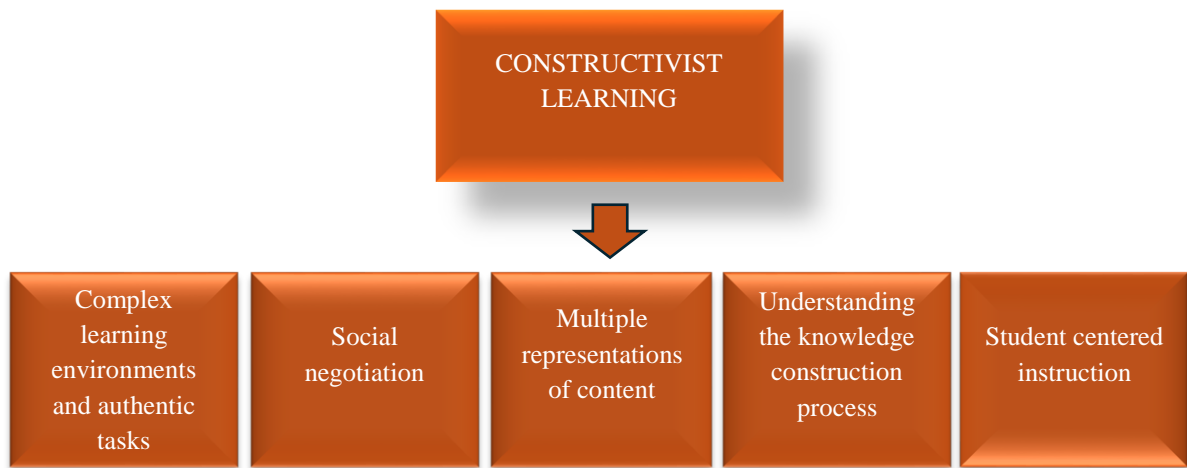


Figure 2.1. Principles of constructivism (Driscoll, 1994; Marshall, 1992).

The theory of constructivism has shaped the instructional practices in EFL classrooms as well. It is maintained that teaching practices in EFL classrooms should address learners' higher-order thinking skills by engaging them in meaningful real-life tasks and should also aim to improve learners' reflective skills through a variety of assessment practices (Abell & Siegel, 2011; Shepard, 2000). In this sense, alternative assessment plays a crucial role in reversing the traditional role of students as passive recipients and offers more opportunities for learner engagement, self-discipline, and autonomy. Furthermore, it fosters the learner's language learning performance, achievement, and language proficiency level as well as improving the learning conditions (Charvade, Jahandar, & Khodabandehlou, 2012).

The constructivist approach has had a huge impact on educational practices in Türkiye as well. This approach has served as a foundation for the new curricula, and the course content, teaching materials, teaching methods, and assessment techniques were all transformed accordingly (Gelbal & Kelecioğlu, 2007). As a result of this reform in the curricula, alternative assessment methods and techniques have gained a wider acceptance in school environments. In particular, the use of projects and other performance-based alternative assessment methods has been suggested along with other traditional assessment techniques (Duban & Küçükylmaz, 2008; Özdemir, 2010). Thanks to these alternative assessment methods, the students are also assessed based on what they can integrate and produce rather than merely on what they can recall or reproduce. In this sense, alternative assessment has a significant place in foreign language classrooms since the main goal of language teaching is to prepare autonomous learners who can function in the real world where all language skills are used in an integrated way.

2.3. Bloom's Taxonomy

Alternative assessment could be also related to Bloom's Taxonomy of Cognitive Objectives. Even though the taxonomy was developed by Benjamin Bloom in the 1950s, it still has a significant impact on teaching practices and assessment policies. Bloom's model encompasses six categories in the cognitive domains, where each domain represents a category. These categories are determined as knowledge, comprehension, application, analysis, synthesis, and evaluation. The names of these categories were modified by Anderson (1999) as it is illustrated in Figure 2.

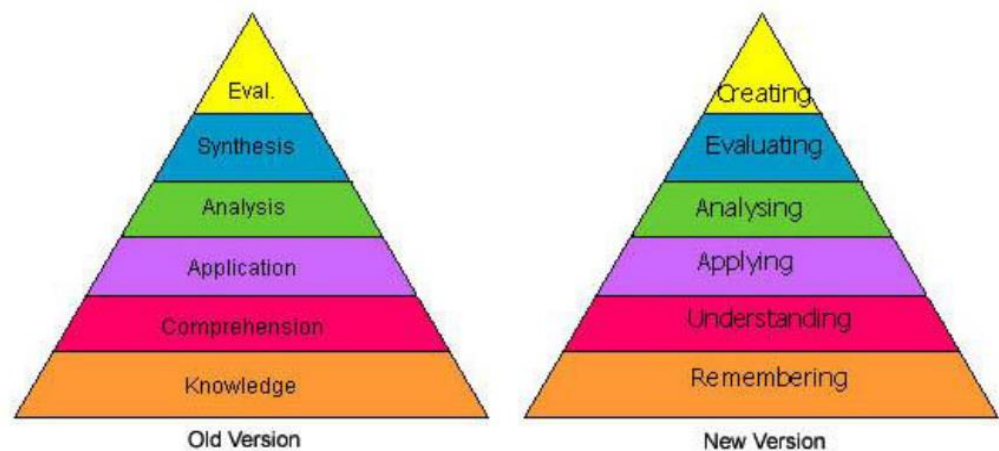


Figure 2.2. Bloom's taxonomy (Bloom, et al., 1956; Anderson, 1999).

It is possible to relate Bloom's Taxonomy to assessment in educational contexts. In traditional assessment, learners are expected to recall information or transfer the knowledge that has been memorized beforehand. Traditional tests generally include test items such as multiple-choice, true-false or gap-filling, and these are related to the very first categories of Bloom's Taxonomy (i.e., remembering and understanding). It is evident that these traditional test items do not serve the purpose of the higher categories such as application, analysis, or evaluation. In fact, traditional assessment methods are criticized since they only address the lower categories of learning in Bloom's Taxonomy and fail to measure analysis, application, and evaluation of the learning (Jacobsen, Eggen, & Kauchak, 2002). Therefore, there is a need for new or alternative assessment tools which can satisfy the higher categories of the taxonomy. Unlike traditional assessment methods, alternative assessment tools and techniques provide learners with an opportunity to master their higher-order skills such as applying, analyzing, creating, evaluating, and creating to attain the learning objectives of the courses.

Bloom`s Taxonomy has had a huge impact on the teaching practices, course materials and assessment applications in EFL classrooms. The recently published coursebooks include specific sections devoted to developing students` critical thinking skills. These sections encourage students to plan and apply what they have learned, analyze, and evaluate the learning process (Anderson, 1999; Bloom et al., 1956). As for the Turkish context, it is evident that foreign language curricula and coursebooks seem to represent the higher categories in Bloom`s Taxonomy. In practice, however, most teachers still focus on the structural aspects of the language and attach great importance to the comprehension of knowledge. This tendency stems from the fact that traditional assessment methods are still common in schools, and this creates a negative washback effect on foreign language education. Therefore, alternative assessment methods have limited space in classrooms, and neither teachers nor students are sufficiently aware of the benefits of such assessment applications. It is therefore important to raise awareness and devote more time and effort to the use of alternative assessment methods and techniques in language teaching.

2.4. Assessment

Assessment is regarded as one of the fundamental components of teaching programs because it influences the attitudes and decisions of several stakeholders such as students, instructors, administrators, parents, and policymakers (Dietel, Herman, & Knuth, 1991). It is a term that has been defined by many scholars in educational contexts. In a broad sense, assessment is described as the process of collecting, recording, interpreting, and making sense of the information about students` performances in certain tasks (Lambert & Lines, 2013). Ming (2002) also defines assessment as the analysis of information concerning the abilities, interests, learning preferences, and achievements of learners. Regarding assessment, Gonzales (2003) points out that it is a systematic process in which teachers monitor their students` performances and provide feedback when necessary. Chen and Hoshower (2003), on the other hand, consider assessment as a pivotal component of learning process, and it could be described as any strategy, tool or technique that can be employed to gather information to track the learners` progress towards the objectives of the course. In a nutshell, it is claimed that assessment can be viewed as an umbrella term which encompasses all kinds of measures that are used to evaluate student achievement (Coombe, Folse, & Hubley, 2007).

According to Rea-Dickens (2000), assessment has several functions in classrooms. To begin with, it acts as a diagnostic tool which provides information to the teacher as to whether students have attained the course objectives. Assessment also helps teachers to identify their students' strengths and weaknesses in certain areas and creates a chance for teachers to reconsider their teaching methods, materials, and assessment tools. Most importantly, it informs school administrations regarding the effectiveness of teaching and learning practices in their institutions. Similarly, Butler and McMunn (2006) state that the purpose of assessment is to provide feedback to both teachers and students so that learning conditions could be improved in the classroom. Finally, Ökten (2009) believes that assessment makes students become more aware of their own learning and helps them to take responsibility in the learning process.

As mentioned above, assessment is of great importance in education, and needs to be handled through appropriate tools to enhance students' learning and develop their skills. In educational contexts, there are various types of assessment methods, each of which serves a different purpose in classrooms. Therefore, teachers should be aware of the strengths and weaknesses of different assessment methods and employ the most appropriate one depending on their objectives of the lesson.

2.4.1. Assessment in English Language Teaching

Testing and assessment are concepts that are closely intertwined in educational institutions and viewed as the cornerstones of any curriculum (Hatipoğlu, 2017). In fact, quality education could only be ensured through valid and reliable assessment practices (Alderson, 2005; Popham, 2006). The recent paradigm shift in language teaching has also changed the function and purpose of assessment. It is commonly acknowledged that learning and assessment need to go hand in hand in classroom settings to attain the objectives of a specific lesson (Davison & Leung, 2009; Stiggins, 2008).

The traditional assessment methods aim to reveal to what extent students have attained the objectives of the lesson through standardized tests. In addition, they are usually applied at the end of the learning process to grade students' performances rather than assessing the learning process. However, due to the emergence of learner-centered language teaching methods, the role and function of assessment has also changed. Nowadays, testing and assessment practices aim to reveal the individual needs of learners in classrooms. In addition, assessment is used as a diagnostic tool throughout the learning process so that teachers could determine what their students lack in certain areas and provide the necessary

knowledge or skills for students (Inbar-Lourie, 2008). Therefore, new, or alternative forms of assessment have emerged in language teaching environments.

The current teaching methods in foreign language education no longer view language learning as accumulation of knowledge or skills. Therefore, the definition and function of assessment has also been modified in recent years. It is now believed that assessment is an inseparable component of the learning process. Regarding this issue, Dann (2002) claims that assessment does not merely include grading students' performances, but also focuses on the process in which learners have an active role in the monitoring of learning and assessment. Additionally, Cheng, Rogers, and Hu (2004) define assessment as "the process of collecting information about a student to aid in decision making about the progress and language development of the student" (p.363). Similarly, Russell and Airasian (2012) view assessment as a tool to gather, synthesize, and interpret information so that certain decisions can be made in classrooms to improve the learning conditions. It is also claimed that assessment has a specific function to support and motivate learners by involving them in decision-making and assessment practices throughout the learning process, which helps them become autonomous learners (Poehner & Inbar-Lourie, 2020).

As can be seen above, the purpose and function of assessment have been altered to focus more on student engagement in the language learning process. In the light of the recent educational reforms, new or alternative assessment tools and practices have emerged to incorporate students into the assessment procedures (Lee & Mak, 2014). By employing alternative assessment methods, teachers become more aware of their students' abilities, identify their strengths and weaknesses, and provide continuous feedback to students to improve their learning (Huang, 2016). In addition, the possible negative consequences of traditional methods, which do not consider the individual needs, interests and learning styles of learners, could be eliminated using alternative assessment methods (Inbar-Lourie, 2008; Shohamy, 2001).

2.5. Assessment Types

In the light of the paradigm shift that occurred in ELT toward learner-centered teaching methods, testing and assessment practices have also been profoundly influenced. There are two main approaches regarding learner assessment in classroom settings (Abbas, 2012; Geberew, 2014). On the one hand, there are traditional assessment methods that are still widely employed in educational institutions due to their practicality

and reliability. On the other hand, there are alternative assessment methods which include a more informal and ongoing process to assess learners' language abilities by requiring them to carry out authentic real-life tasks through the integration of skills (Agustina, 2011).

2.5.1. Traditional Assessment

Traditional assessment tools are commonly administered to assess what students can recall or recognize based on the knowledge that was transferred by the teacher to the students during the lesson (Serafini, 2001). These assessment tools are regarded as traditional or conventional because they tend to be indirect, inauthentic, standardized, and formal (Benzehaf, 2017; Coombe & Hubley, 2011). They generally include select-response test items such as multiple-choice tests, fill-in-the-blank, true/false, and matching exercises (Belle, 1999, as cited in Wikström, 2008; Brown & Hudson, 1998). In such tests, students are generally presented with some question items and expected to select the correct answer from a certain set of options provided to them. As one can see, no language production is required from students in these tests (Brown & Hudson, 1998). According to Brown (2004), the purpose of traditional assessment methods is to reveal what students have learned at the end of the teaching process. In other words, students' performance is assessed in a summative manner to see whether the learning objectives of the lesson have been accomplished or not by the students. At the end of the summative assessment conducted through traditional tools, students' performances are usually reported with certain grades indicating whether they have passed certain tests. In addition, the results of such tests could be used to make certain decisions at the end of the teaching process (Irons, 2008; Mede & Atay, 2017).

On the other hand, it is asserted that there are legitimate reasons for the use of traditional assessment methods. To begin with, traditional assessment techniques and methods are still widely used as effective tools in classroom settings since high-stakes exams are also designed with a traditional approach in many countries around the world. It is claimed that using traditional methods facilitates test administration, ensures scoring objectivity, and increases score reliability since certain devices are used for scoring (Brown & Hudson, 1998; Simkin & Kuechler, 2005). Therefore, proponents of traditional assessment methods assert that using such methods is the most convenient way of assessment which allows precise identification and monitoring of student progress (Kwako, 2003). It is possible to employ such methods in large-scale examinations due to their low cost of preparation. In brief, governments, educational institutions, and teachers opt for traditional

assessment methods as they are as practical, reliable, valid, and economical though they are known to have several shortcomings.

2.5.1.1. Criticism of traditional assessment. Even though traditional assessment methods are considered to be practical due to the above-mentioned reasons, many scholars and educationalists have reported several weaknesses regarding the use of these methods. Proponents of the process-oriented teaching and assessment methods regard traditional assessment ill-suited for foreign language classes since they cannot capture learners' actual progress in class or identify their strengths and weaknesses in certain areas (Balliro, 1993; Barootchi & Keshavarz, 2002; Williams, 2008). In addition, Braun and Mislevy (2005) state that traditional assessment tools measure learners' ability to recognize or recall information and cannot address higher order thinking skills of learners. Therefore, it leads to short-term, superficial learning and only caters for the lower level of Bloom's taxonomy (Simonson, Smaldino, Albright & Zvacek, 2009). Similarly, Çakır (2013) states that it is almost impossible to measure learners' higher order thinking skills, their problem-solving abilities and underlying knowledge through traditional assessment tools. Finally, due to their test item formats, these tools are unable to reveal whether learners' have acquired the 21st-century competencies such as critical thinking and problem-solving (Koh, 2017).

According to Bailey (1998), traditional assessment methods are generally indirect and inauthentic since they include decontextualized test items in a discrete form. Regarding this issue, Brown and Hudson (1998) stress that such methods fail to represent real-life language and only expect students to provide answers for discrete test items. Indeed, Lizasoain and Zárata (2014) maintain that traditional assessment methods generally assess language performance rather than learners' competence. Nevertheless, learners' performance in a language does not necessarily reveal their underlying knowledge in that language. To support this view, Nasab (2015) asserts that traditional assessment methods generally focus on students' receptive skills and ignore the productive language skills of learners. In fact, traditional assessment methods generally do not involve classroom projects, discussions or other performance-based activities in which students can demonstrate their individual skills (Franklin, 2002). Therefore, if these methods are used as the sole indicators of students' abilities or achievements, they may generate inaccurate results (Barootchi & Keshavarz, 2002).

Traditional assessment methods are also believed to have a negative washback effect in EFL classrooms (Hughes, 1990). Since the main goal of teaching is to pass certain tests

which include multiple-choice, gap filling or true/false test items, both teachers and students tend to neglect productive language skills as these tests direct teachers to focus on only those subjects that are included in the examinations (Shepard & Dougherty, 1991). In other words, they have a huge impact on the language curriculum and teaching methods which attach great importance to exam preparation practices at the expense of real-life authentic activities (Wall, 1996).

Most importantly, traditional assessment might cause stress and anxiety for students since they include norm-referenced speed-based tests in a summative form. Students struggle to get high grades on these tests since they are regarded as the sole indicators of achievement for teachers, students, and their parents as well (Bailey, 1998). Therefore, traditional assessment methods create a competitive learning environment for students, which might have adverse effects on students' attitudes and feelings towards learning. In other words, at the psychological level, students are affected negatively by these methods as their interests, motivation and efforts are neglected in the classroom (Broadfoot, 2003). In addition, these standardized tests can prevent learners from expressing their knowledge in personal terms and hinder creativity in classroom, which leads to boredom and a lack of interest among learners (Simkin & Kuechler, 2005; Underhill, 1992). Finally, these tests promote extrinsic motivation for students since their main goal is to pass exams rather than learning the language for real-life purposes (Brown & Abeywickrama, 2010).

Due to the abovementioned reasons, it is obvious that using traditional, summative testing in EFL classrooms is no longer sufficient for effective language teaching. As a result of this dissatisfaction with traditional assessment methods, scholars who are interested in assessment methods and tools coined the term "alternative assessment," which indicates a radical shift from traditional assessment methods towards authentic, direct, and process-based assessment applications. The following table illustrates this paradigm shift in assessment.

Table 2.1. Paradigm Shift from Traditional Assessment to Alternative Assessment

Previous Paradigm/TA	Current Paradigm/AA
Focus on language	Focus on communication
Teacher-centered	Learner-centered
Isolated skills	Integrated Skills
Emphasis on product	Emphasis on process
One answer, one-way correctness	Open-ended, multiple answers
Tests-oriented	Learning-oriented

Adapted from Johannessen & Redecker (2013) and Letina (2015).

In the following section, the definitions, and key characteristics of alternative assessment, which is also the focus of this study, are presented in detail.

2.5.2. Alternative Assessment

As a result of the increasing criticism of traditional assessment methods, a need has emerged for new or alternative assessment methods that include authentic, real-world tasks (Brown, 2004). In literature, several definitions have been proposed by scholars regarding alternative assessment. To start with, Hancock (1994) defines alternative assessment as an ongoing teaching and learning process that involves both students and teachers in decision-making about the learning process through non-conventional tools. McNamara (2001), in a similar vein, views alternative assessment as a movement “away from the use of standardized multiple-choice tests in favor of more complex performance-based assessments” (p.329). Additionally, Richards and Renandya (2002) propose alternative assessment as a complement to traditional testing tools since there are several shortcomings of these tools.

Barootchi and Keshavarz (2002) point out that alternative assessment is regarded as non-traditional assessment which encompasses any types of assessment apart from standardized, traditional tests. In literature, scholars have used several terms interchangeably for alternative assessment. The most common terms used by scholars are authentic assessment, performance assessment, direct assessment, and informal assessment (Hamayan, 1995; Herman, Aschbacher & Winters, 1992). According to Hughes (2003), alternative assessment can also be called as a performance-based classroom assessment. These include several activities such as portfolio, project, presentation, role play, self-assessment, peer assessment and learning journal (Brown, 2004; Greenstein, 2010; Richards & Schmidt, 2002).

2.5.2.1. Benefits of alternative assessment. Scholars and educationalists have been trying to come up with better ways to assess students’ language skills and knowledge on the grounds that traditional assessment methods fail to capture an accurate picture of their real competences. To fill this gap, alternative assessment methods have gained a wider acceptance in EFL classrooms recently since they are believed to present new ways of measuring students’ language ability through authentic real-life tasks (Reeves 2000, as cited in Nasab, 2015) The concept of authenticity creates an opportunity for students to be involved in authentic tasks that reflect real-life situations.

In this vein, Brown and Hudson (1998) have listed twelve key characteristics of alternative assessments as follows. Alternative assessment methods;

1. require students to perform, create, produce, or do something;
2. use real-world contexts or simulations;
3. are nonintrusive in that they extend the day-to-day classroom activities;
4. allow students to be assessed on what they normally do in class every day;
5. use tasks that represent meaningful instructional activities;
6. focus on processes as well as products;
7. tap into higher-level thinking and problem-solving skills;
8. provide information about both the strengths and weaknesses of students;
9. are multiculturally sensitive when properly administered;
10. ensure that people, not machines, do the scoring, using human judgment;
11. encourage open disclosure of standards and rating criteria; and
12. call upon teachers to perform new instructional and assessment roles. (p.654-655).

Scholars have also suggested that using alternative assessment tools provides several advantages both for teachers and students. To begin with, it can provide valuable information for teachers about their students' performance in classroom settings (Barootchi & Keshavarz, 2002). In addition, alternative assessment "connects students' experiences with the curriculum through active involvement" and thus promotes meaningful learning and in-depth teaching (Gottlieb, 2006, p.111). As a result, it creates a positive washback effect on learning (Norris, Brown, Hudson & Yoshioka, 1998).

Regarding the benefits of alternative assessment methods, Jacobs and Farrell (2001) claim that alternative assessment tools reflect real-life situations and improve learners' thinking skills. To support this claim, they emphasize three features of alternative assessment in EFL settings. First, alternative assessment methods focus on meaning rather than form. Second, they attach importance to the learning process as well as the product. Third, alternative assessment provides a social, collaborative learning environment through peer assessment and group activities.

Another positive aspect of alternative assessment is that this type of assessment is based on a collaborative approach which creates an opportunity for interaction between students and the teacher during the learning process (Barootchi & Keshvarz, 2002). Therefore, a rapport is built in the classroom between students and the teacher. This positive atmosphere in turn improves students' self-esteem and boosts their intrinsic motivation and sense of self-efficacy (Broadfoot, 2003). In brief, alternative assessment has a positive effect

on the motivation and attitudes of both students and teachers throughout the learning process (Karim, Amir, Darman & Muhammad, 2018).

Finally, alternative assessment is regarded as an effective tool to promote the intellectual development of learners as it requires them to demonstrate their problem-solving skills, higher order thinking and creativity in real-world tasks using their underlying knowledge (Brawley, 2009; Frey & Schmitt, 2007). While traditional methods dictate learners the existence of only one correct answer, alternative methods allow them to explore alternative possibilities, enabling them to develop their 21st-century skills and competencies in EFL classrooms (Koh, 2017). Table 2.2. highlights the distinctive features of both assessment types and summarizes the differences between traditional and alternative assessment methods (Brown, 2004).

Table 2.2. *Traditional and Alternative Assessment*

Traditional Assessment	Alternative Assessment
One-shot, standardized exams	Continuous, long-term assessment
Timed, multiple-choice format	Untimed, free response format
Decontextualized test items	Contextualized communicative tasks
Scores suffice for feedback	Individualized feedback and washback
Norm-referenced scores	Criterion-referenced scores
Focus on the right answer	Open-ended creative answers
Summative	Formative
Oriented to product	Oriented to process
Non-interactive process	Interactive process
Fosters extrinsic motivation	Fosters intrinsic motivation

Adapted from Brown (2004, p.13).

According to the table provided above, it is obvious that the traditional assessment methods have several deficiencies in the learning process. On the other hand, the alternative assessment methods compensate for the shortcomings of traditional assessment methods, and provide a more accurate, fair, and effective measure of students' performances. However, it is essential to know which alternative assessment tools are suitable for a specific group of students because there are various types, and each serves a different purpose in learning.

2.5.2.2. Challenges of alternative assessment. Despite the abovementioned benefits of alternative assessment applications, scholars and teachers have stated a number of concerns and limitations regarding the use of alternative assessment methods and tools in classrooms. To start with, teachers could have some problems in applying alternative

assessment methods. For instance, designing real-life tasks can be challenging for some teachers (Leung & Lewkowicz, 2006). These tasks are supposed to match the objectives of the curriculum and reflect the real performance of learners. However, some teachers may lack the skills or expertise to design appropriate tasks, manage the assessment process and evaluate learners' performances in a valid and reliable way (Fuchs, 1995). Another difficulty of implementing alternative assessment is setting appropriate criteria for measuring and evaluating learners' performances. According to Maclellan (2004), it can be really challenging for some teachers to design rubrics, checklists or other assessment tools which can be used to measure learners' performances in a valid and reliable way. In addition, there are usually more than one assessor in the assessment process, so it takes a lot of time to agree on the final grades of learners, especially in portfolio tasks (Song & August, 2002). As one can see, some teachers generally have inadequate experience in designing alternative assessment tasks and scoring rubrics due to lack of knowledge and expertise. Therefore, they require continuous training and professional development opportunities from their administrators, which could be time consuming and costly (Demir, 2021). Even if these teachers have received training on how to implement alternative assessment methods, they may still stick to their traditional ways of assessment due to exam pressure and curricular intensity. Therefore, it is not surprising that teachers tend to avoid using alternative assessment in their classrooms.

Alternative assessment methods may also not be favored by some students. As most students are concerned about their performances in standardized high-stakes exams, they might resist the use of alternative assessment methods which require learners to undertake new roles and responsibilities in classrooms (Tedick & Klee, 1998). In traditional classrooms, students generally focus on learning information and rules to pass certain tests. In other words, they are passive recipients of knowledge. On the other hand, alternative assessment methods engage students both in the learning and assessment process actively. Therefore, students may be asked to carry out a wide range of authentic real-life tasks that require a lot of preparation and collaboration both inside and outside of the classroom. However, due to exam pressure, some students might be reluctant to participate in such tasks. In alternative assessment, students could also have active roles during the assessment process. For instance, they might be asked to assess their peers' performances based on certain criteria and provide feedback for them. However, most students view the teacher as the sole authority in traditional classrooms (Zhang & Hyland, 2018). Therefore, they might be skeptical that their peers could contribute to their learning through feedback and thus

prefer to receive feedback from the teacher (Fathi, Afzali, & Parsa, 2021). In short, it can be difficult to encourage students to assume new roles and undertake new responsibilities during the alternative assessment applications.

Finally, alternative assessment might require a policy change at institutional level since each institution has its own procedures and regulations regarding assessment (Demir, 2021). However, it can be difficult to adopt alternative assessment methods due to crowded classrooms, lack of teaching staff and other resources. Therefore, many institutions still apply traditional sit-down exams in their classrooms due to their practicality.

2.5.2.3. Solutions for the challenges. As can be seen above, there are a number of challenges in the implementation of alternative assessment methods. However, scholars have suggested several solutions to deal with these challenges. Regarding validity and reliability concerns, Huerta-Macias (1995) argues that as alternative assessment measures students' performances through authentic tasks such as role-playing, participating in pair or group works, presenting in front of others etc.; these procedures reflect the real-life situations. In addition, students are supposed to use four language skills in an integrated way to complete alternative assessment tasks, so teachers can capture a more holistic picture of students' language performances. Therefore, alternative assessment is considered valid in many respects.

To evaluate the validity of alternative assessments, Baker (2010) suggested two main criteria, which are related to *internal validity* and *external validity*. To meet the criterion of internal validity, alternative assessment needs to include cognitively complex tasks, and students need to be able to achieve "the intellectual demands of the tasks" (Baker, 2010, p. 15). Therefore, teachers need to make sure that these tasks are not based on memorization of knowledge or rules, which reflects the superficial features of assessment. Another feature of internal validity is related to the meaningfulness of the tasks. This means that the purpose and components of these tasks should be comprehensible to all students to avoid ambiguity or chaos so that they can feel motivated to perform them. In addition, students should be provided with clear guidance and support to complete these tasks. Alternative assessment tasks should also have quality to meet the criterion of internal validity. This means that the tasks need to be designed in parallel with the objectives of the lesson. Finally, alternative assessment outcomes ought to be generalized and transferred to other topics or domains of learning (Baker, 2010). For instance, if a student has learned how to present his/her ideas in a specific topic, he/she is expected to do the same in other subjects or topics as well.

As for external validity, the first criterion is related to the consequences of alternative assessment. According to Messick (1989), any type of assessment is evaluated in relation to its consequences. Therefore, alternative assessment is expected to yield positive consequences as long as it enhances student interest and motivation in learning. Another consequence of alternative assessment could be observed in teaching strategies. When teachers recognize the value of alternative assessment, they could adopt new teaching strategies or redesign their teaching materials to accomplish the objectives of the lesson. For instance, if a teacher aims to improve the problem-solving skills of learners using alternative assessment methods, this is considered a positive consequence of assessment (Cunningham, 1998). Another important criterion of external validity is related to the fairness of assessment. This means that all learners should be given equal support, guidance, and time to be prepared for assessment. In addition, their performances need to be assessed through standardized rubrics to avoid ambiguity or bias. Finally, the alternative assessment ought to be evaluated in terms of cost and efficiency. Stakeholders could allocate a great amount of budget to design or evaluate alternative assessment tasks. For instance, they might buy some technological devices or software programs to implement alternative assessment methods. In addition, both teachers and students spend a great deal of time and effort during the alternative assessment process. Therefore, the implementation of alternative assessment should be evaluated carefully in terms of its cost and effectiveness.

As for the grading of students' performances in alternative assessment applications, the issues of objectivity, reliability and validity pose a great challenge to all stakeholders. Baker (2010) suggests a few solutions to deal with this challenge. First of all, the use of rubrics needs to be encouraged in classrooms since they contain transparent scoring criteria both for teachers and students. If teachers receive expert views and design appropriate rubrics based on the objectives of the lessons, they could conduct a fair, reliable, and valid assessment in the classroom. Secondly, interrater reliability of this assessment could be ensured with the help of multiple assessors. In this way, trust is built between the teacher and students in terms of the accuracy of assessment. Finally, students need to be informed that both the process and product are evaluated in alternative assessment, and this should be reflected in the rubrics as well. If all these issues are formalized as a part of assessment procedures, consistency could be ensured in all classrooms (Moskal & Leydens, 2000).

In addition to the issues mentioned above, there are some other factors that need to be considered while implementing alternative assessment in EFL classrooms. First of all, many teachers complain that they cannot apply alternative assessment methods due to a lack

of expertise in this field. Therefore, administrators need to provide training for teachers on how to implement alternative assessment in classrooms. In addition, workshops, seminars, or other in-service courses could be organized to create awareness on the benefits of alternative assessment. Another major factor that prevents teachers from implementing alternative assessment is exam pressure. Since there are many traditional sit-down exams during the year, teachers struggle to cover all the exam subjects in the course syllabi. As a result, they are unable to apply alternative assessment methods in their classrooms. Nevertheless, teachers could still apply alternative assessment as a low-stake assessment practice and integrate this assessment into their overall grading criteria (Demir, 2021).

2.6. Types of Alternative Assessment

Since the 1990's, there has been a growing interest in the implementation of new or alternative assessment methods since traditional assessment tools have merely focused on improving learners' receptive skills rather than promoting communicative competence (Klenowsky, 2002, as cited in Bahous, 2008; Maclellan, 2004). Furthermore, there was not much interaction or collaboration in classroom due to these teacher-centered traditional assessment methods. Therefore, educationalists and scholars have looked for innovative ways of assessment that aim to increase students' engagement in the learning process (Bahous, 2008). In the relevant literature, these new ways of assessment have been termed in different ways such as alternative assessment, performance assessment, authentic assessment, direct assessment, and so on. To be consistent and avoid ambiguity, alternative assessment will be used in this study. Hamp-Lyons (1992) made a distinction between the activities that yield data for assessment and the ways through which teachers organize and record the data. In parallel with this view, Reeves (2000) states that alternative assessment procedures generally involve two main approaches. These are portfolio assessment, and performance or authentic assessments which include debates, role plays and presentations (Reeves, 2000). The figure below shows the most commonly employed alternative assessment tools and methods in educational settings.

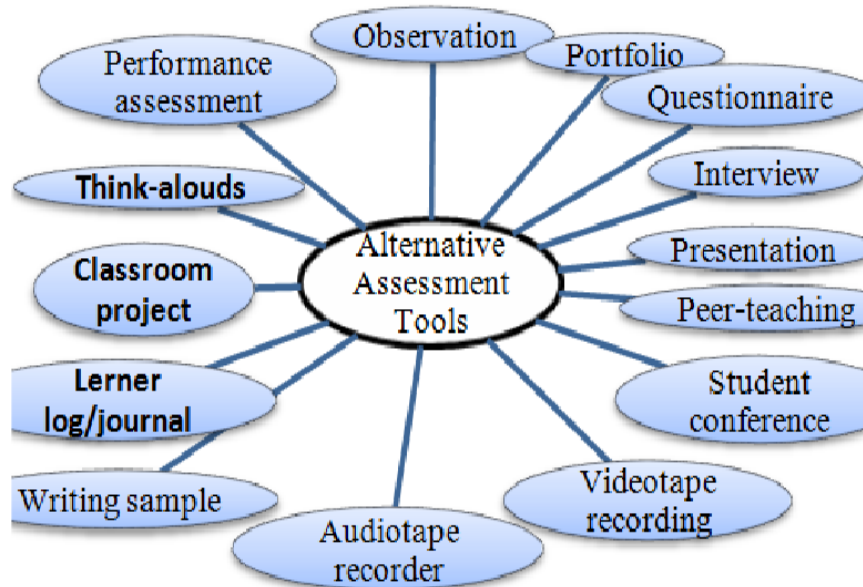


Figure 2.3. Alternative assessment tools and methods. Adapted from Herdiawan (2018) and Nasab (2015).

In the following section, definitions and features of the alternative assessment procedures employed in this study have been presented.

2.6.1. Portfolios

Portfolio is regarded as one of the most prominent and commonly employed alternative assessment methods (Burnaz, 2011). In the literature, a number of scholars have defined portfolio in several ways. According to Paulson and Meyer (1991), it is a purposeful and systematic collection of students' works which also reflect their efforts and progress in a specific area. Similarly, Wolf and Siu-Runyan (1996) describe portfolio as "a selective collection of student work and records of progress gathered across diverse contexts over time, framed by reflection and enriched through collaboration, that has as its aim the advancement of student learning" (p. 31). Some scholars in ELT field have described portfolio assessment by focusing on writing skill specifically. According to Hyland (2003), portfolio assessment can be viewed as a reaction to traditional, one-shot tests where learners are not given the opportunity to revise, redraft, and choose their best works for grading. In their definition of portfolio, Coombe and Barlow (2004) states that "as far as portfolios are defined in writing assessment, a portfolio is a purposive collection of student writing over time that shows the stages in the writing process a text has gone through and the stages of the writer's growth" (p. 19).

While the definitions presented above might reflect different perspectives, they all highlight the idea that portfolios are mostly used to monitor and improve students' progress throughout the learning process. In this sense, they are regarded as one of the most prominent alternative assessment tools used in foreign language education (Hamp-Lyons, 1995). It is argued that portfolio assessment serves the "need to assess more complex phenomena," and they are generally referred to as one of the best examples of authentic or alternative assessment (Hamp-Lyons & Condon 2000, p.18). Similarly, Chang and Tseng (2009) point out that portfolios are commonly employed in tertiary education as an authentic assessment tool that shows learners' real performances throughout the learning process.

Regarding the content of portfolios, Brown (2000) suggests that they can include a wide range of students' products such as audio or video recordings, essays, research projects, stories, and any artistic works. Apart from the content of a portfolio, there are some other issues that need to be considered carefully. First, students need to be provided a guideline that demonstrates the purpose, content, and assessment criteria of portfolios. Second, they should be introduced a sample of portfolio work so that students get familiar with portfolio practice and have some ideas about what they are expected to do throughout the process. Finally, a schedule which shows important dates and deadlines regarding the portfolio tasks should be shared beforehand so that students can manage their time and effort accordingly (Brown, 2000).

As an alternative assessment method, portfolios have been extensively used in foreign language classroom settings, and previous research has revealed several advantages of portfolio practice (Burnaz, 2011; Fox, 2017). It is argued that portfolios encourage learner involvement in the assessment procedures, and make them responsible for their learning, which creates a sense of ownership of the learning process. In other words, it helps students to become autonomous learners (Genesee & Upshur, 1996; Hirvela & Pierson, 2000). In addition, Brown and Hudson (1998) claim that portfolio assessment helps students feel motivated towards learning since it allows interaction and collaboration in the classroom. In this way, students can get support and feedback from both their peers and teachers and learn better (Hancock, 1994). Therefore, portfolio implementation focuses on the selective, reflective, and collaborative collection of evidence regarding the learner's performance (Chirimbu, 2013).

Another benefit of portfolio keeping is that it provides an ongoing, longitudinal performance assessment while traditional tests are applied within a limited time, which

creates pressure for students (Bailey, 1998). Regarding this issue, Hamp-Lyons and Condon (2000) suggest that portfolios are particularly useful for EFL learners since they provide a broader assessment of what these learners can accomplish within a more flexible time context. According to Richards and Renandya (2002), portfolios provide evidence of students' progress from the start to the end of the learning process, giving both the teacher and students an opportunity to reflect on how much has been achieved. By looking at their portfolios, students will be informed about their own progress, and take some responsibility for the assessment of their learning and become more aware of their abilities, strengths, and weaknesses in different skills. (Aydın & Başöz, 2010; McMillan, 2001; Salkind, 2006). In other words, they become autonomous learners (Richards & Renandya, 2002).

Portfolios also provide several benefits for teachers regarding their instructional practices. First, portfolios can reveal different learning styles and meet individual needs of learners so that teachers can adjust their teaching accordingly. Another important advantage of using portfolios in foreign language classrooms is that it helps teachers to understand students' ideas and feelings regarding the learning process through self-reflection. This creates a chance to improve dialogue and collaboration between the teacher and students as well as building a rapport in the classroom (Mullin, 1998; O'Malley & Pierce, 1996; Salkind, 2006; Valencia & Calfee, 1991). Finally, teachers could get a more accurate picture of their students' overall performances through portfolios since they include a wide range of data regarding their language skills.

In short, the abovementioned citations indicate that using portfolio as an alternative assessment tool provides numerous benefits for both teachers and students. When implemented in a systematic way, portfolios can enhance students' performances and increase their motivation towards learning in EFL classrooms. Furthermore, teachers can gather rich data regarding the effectiveness of their instructional practices and make the necessary adjustments when necessary.

2.6.2. Performance Assessment

Performance assessment is another alternative assessment method that includes oral presentations, role-plays, debates, and other tasks that require learners to produce. In a performance task, students are expected to create, produce, implement, and present their ideas regarding real-life problems. Oral presentations are widely employed as a performance assessment method to evaluate learners' proficiency levels or specific language skills. This type of assessment requires learners to do search and collect information on specific issues

and present their findings in classrooms. In this sense, a presentation incorporates learners' productive and receptive language skills in an integrative manner in the context of a communication-oriented approach (Herdiawan, 2018).

Role-play is another commonly used activity to promote communication and interaction in EFL classrooms. Tompkins (2001) describes role-play as an activity that encourages learners to participate in the learning process actively by taking on certain roles in specific contexts. Since many learners do not have a chance to use the target language in real-life situations, they are provided with this opportunity through meaningful role-play activities. In other words, they have a chance to practice the target language in authentic situations without feeling anxious or stressed (Kenneth, 2008). In addition, role-plays bring variety and entertainment to EFL classrooms. Therefore, they have a beneficial effect on learners' attitudes towards learning as well as improving their oral skills (Cornett, 1999).

Debate activities are also used in EFL classrooms to foster students' communicative competence. For instance, Akerman and Neale (2011) describe debates as communicative events in which a minimum of two groups with opposing views follow a set of pre-determined rules to convey their points of views on a specific issue. Students are usually assigned debate topics and given sufficient time to prepare for the debate time. During a debate, each group struggles to defend its ideas and respond to the opposing views of the other group. Therefore, there is a great deal of collaboration and interaction during a debate activity. Debates are regarded as one of the most effective alternative assessment activities since they promote linguistic, social, and cognitive development of learners through social interaction (Firth & Wagner, 2007).

As an alternative assessment method, performance assessment includes various activities in which learners actively participate in the learning process. For an effective performance assessment, two things need to be considered carefully. The first one is to establish suitable tasks which serve the objectives of the lessons. The second one is to inform learners about the assessment criteria so that transparent scoring can be ensured. Finally, it is essential to provide constructive feedback to the learners related to their performance (Backman & Klinghammer, 2006).

2.6.3. Learning/Reflective Journals

In recent years, the use of journals as an alternative assessment method has become prominent in EFL classrooms. Learning journals, also known as reflective journals, creates a space for learners so that they can convey their views and feelings freely towards classroom

practices (Brown, 2000). It is asserted that the practice of keeping learning journals not only reinforces learning but also promotes critical thinking (Moon, 2004; Tsang, 2003). Moon (1999, as cited in Prinsloo et al., 2011) states that learning journals have various purposes. To begin with, they enable students to record their learning experience, and give them a chance to reflect on their learning process. Regular entries students make in their journals provide teachers with an insight into their students' progress in learning as well as their perceptions of the learning process. In addition, the use of learning diaries encourages students to reveal their true feelings and attitudes towards classroom practices, course materials, and examinations. In this sense, it promotes self-expression among students, and their voices are heard by the teacher to improve the learning conditions in the classroom (Moon, 2004). As mentioned above, learning journals serve several purposes in classroom settings and are generally employed as an alternative assessment tool in EFL classrooms due to their benefits. However, as in any activity, the real purpose of keeping journals needs to be explained clearly to the learners so that they can reflect their ideas and feelings freely through their entries.

2.6.4. Conferences

Conferences are defined as informal one-to-one conversations between teachers and students (Fenner, 2013). The main purpose of a conference is to assess learners' work samples using predetermined criteria and provide constructive feedback to improve them. Conferences also create an opportunity for negotiation between learners and teachers, allowing them to clarify ambiguities, set goals, and share expectations related to the learning process (Nasab, 2015). As an alternative assessment method, conferencing is widely used in EFL classrooms as a response to students' performances, especially in writing and speaking lessons. In the present study, the researcher had one-to-one conferences with the participants to provide feedback on their performances and discuss their strengths and weaknesses after each performance task.

2.6.5. Self-Assessment

In educational settings, self-assessment is widely used as an alternative assessment method, especially in EFL classrooms. According to Hargreaves and Fullan (1998), self-assessment can be defined as learners evaluating and judging the quality of their products using predetermined criteria. Regarding self-assessment, Backman and Klinghammer (2006) state that it is considered a reflective process in which students have a chance to assess their

performances based on certain criteria or guidelines that have been determined and shared with the students in advance.

Scholars and researchers have identified several benefits of self-assessment. In this vein, Brown and Hudson (1998) state that using-self assessment fosters learner autonomy and increases learner engagement in classrooms. According to McDonald and Boud (2003), self-assessment not only encourages students to reflect on their learning but also makes them aware of their strengths and weaknesses in specific areas. Therefore, it creates an opportunity for students to improve their language skills on their own, making them independent, self-critical, and self-reliant learners (Janisch, Liu & Akrofi, 2007). Self- assessment is especially beneficial in improving the writing skills of learners. As students learn how to evaluate their writings, based on certain criteria, they tend to become proficient writers in time.

Self-assessment also has a positive effect on students' attitudes towards learning and assessment since it reduces the pressure of teacher assessment and empowers students in the assessment process. In addition, if students know how to evaluate their own performances, they are more likely to become confident in their abilities and accomplish difficult tasks (Kastrati, 2013). Self-assessment is a useful tool for teachers as well. Traditional assessment methods provide little or no information regarding students' states of mind during the learning process. In contrast, self-assessment techniques reveal crucial data related to learners' efforts, strengths, and weaknesses as well as their expectations. As a result, teachers have a better understanding of their students' needs and take action to address them.

Nowadays self-assessment is viewed as an indispensable part of the learning process and its effect on learners' performance has been highlighted by many scholars (McMillan & Hearn, 2008; Ross, 2006). However, there are certain conditions which need to be ensured so that self-assessment could be applied successfully. First, students need to be provided information on this alternative method and be aware of the purpose and goals of this assessment process. In addition, the implementation process should be strictly controlled and the criteria, if used for score giving, should be explicit, clear, and introduced beforehand. Finally, quantitative results need to be supplemented by qualitative methods (e.g., conferences) because only in this way will self-assessment alternative assessment be beneficial for both teachers and students. In the present study, the participants were engaged in self-assessment in their writing lessons using the rubric selected for the writing lesson as well as the checklists in their writing coursebooks.

2.6.6. Peer Assessment

To understand peer assessment, it is necessary to mention Vygotsky's zone of proximal development (ZPD henceforth). Vygotsky (1987) defines ZPD as the distance between what learners can do independently and what they are capable of doing in collaboration with more competent peers. Therefore, Vygotsky (1986) suggests that peer assistance is of great significance in learning, and both the provider and the receiver of peer feedback benefit from this mutual scaffolding process.

In EFL classrooms, peer assessment is widely used as an alternative assessment method. Topping (2009) defines peer-assessment as a process in which learners provide constructive feedback and support one another to improve their learning. Peer assessment is highly favored in student-centered classes as this process promotes collaboration and interaction among students. In addition, it fosters learners' critical thinking skills, and allows for reflection through individualized feedback sessions (Davies, 2006). In other words, peer assessment provides learners with the opportunity to monitor both their own and peers' learning processes, making them aware of their capabilities, strengths, and weaknesses in the process. If this peer assessment process is managed properly, it can make students autonomous learners by reducing dependency on teacher assessment.

According to Cheng and Warren (2005), peer assessment has been mostly incorporated into EFL writing classes where students provide feedback for each other's written works to improve them. Therefore, it creates an opportunity for students to receive more feedback on their writings and gain new perspectives through meaningful interaction with their peers. On the other hand, like self-assessment, peer assessment also requires clear guidelines and training to be effective for students (Min, 2006). In addition, teachers need to observe how students grade their peer's papers and reserve the right to make any adjustments when necessary (Roberts, 2006). Since most students view the teacher as the sole authority, they might regard teacher assessment to be more accurate and reliable (Zhang & Hyland, 2018). Therefore, a negotiation needs to be achieved between the teacher and students on their final grades. In the present study, the participants were engaged in peer assessment using the rubric selected for the writing lesson as well as the checklists in their writing coursebooks.

2.7. Tools for Alternative Assessment

In EFL classrooms, the outcome of any alternative assessment activity or performance is documented and evaluated through certain assessment tools. These tools are designed based on the objectives of the assessment. While most teachers adopt a holistic approach to assess their students' performances, it is also possible to analyze and evaluate specific aspects of learners' products (Harp, 1991; Hill & Ruptic, 1994). The following tools are widely used in foreign language classrooms as a means of documenting and evaluating learners' performances.

2.7.1. Grading Rubrics and Checklists

Linn and Miller (2005) define a grading rubric as a set of standards or guidelines through which students' performances are assessed. Rubrics are mostly used to assess the written or oral products of learners. Higgins (2011) discusses two main categories of rubrics which are holistic rubrics and analytic rubrics. Holistic rubrics are employed to obtain a general assessment of a learner's product while analytic rubrics demonstrate a wide range of discrete indicators of achievements. A grading rubric typically includes categories ranging from advanced to beginning levels of students' performance. Regarding this issue, Airasian and Russel (2008) point out that grading rubrics are highly effective when they are used by teachers who aim to focus on multiple dimensions in a single written or oral product to identify their students' strengths and weaknesses. When a teacher considers implementing alternative assessment, s/he can also design a checklist for evaluation. A checklist is defined as a written description of dimensions that must be observed in an acceptable performance of a student. In alternative assessment applications, it is essential to inform students about the scoring criteria in advance so that they will be aware of what their teacher is looking for in their products. To do this, grading rubrics and checklists are introduced to students at the very beginning of the learning process, and grading transparency is ensured by the teacher (Linn & Miller, 2005). In the literature, practitioners have been offered three options while deciding on a rubric. They could use an already existing rubric without any changes, make some modifications on them to suit their purposes or create an entirely new one (Perlman, 2003). In this study, the researcher, and his supervisor made some modifications on the rubrics that they gathered in the literature and online sources. Their purpose was to gather detailed information on the participants' performance on specific alternative assessment tasks.

2.8. Previous Studies on Alternative Assessment

Most of the EFL studies conducted on alternative assessment have focused on two main issues. While some studies have investigated the effect of applying certain alternative assessment methods on learners' performances in EFL learning, others have attempted to find out the effects of alternative assessment methods on students' perceptions, motivation levels or how these assessment methods affected their attitudes towards learning English. The following sections will present the findings of some of the recent studies conducted on alternative assessment.

2.8.1. Studies Conducted on Students' Academic Achievement

2.8.1.1. Portfolio studies. In one of the earliest studies, Yurdabakan and Erdogan (2009) attempted to investigate the effects of portfolio assessment, a prominent alternative assessment method, on the listening, reading, and writing skills of a group of secondary school preparatory class students in Türkiye. The study also explored the views of the participants regarding the use of portfolio assessment. The participants of the study included 44 high school students who were placed in treatment and control groups equally. The researchers claimed that the study was unique since it not only addressed writing skills but also listening and reading skills as well. Prior to the treatment process, a pre-test was administered to both groups to determine their existing knowledge of each skill. During the treatment process, which lasted 12 weeks, the participants in the experimental group were involved in portfolio assessment applications in three skills while the participants in the control group were taught in a traditional way. In other words, the lesson materials and assessment tools were totally different from each other in both groups. After the treatment process, a post-test was given to both groups to see the effects of two different implementations. In addition, the participants in the experimental group were asked to respond to six open-ended questions which aimed to reveal their opinions on the portfolio assessment process. The results of the statistical analyses indicated that there was not a significant difference between the two groups in terms of their listening and reading scores. However, it was found that the participants in the experimental group outperformed the participants in the control group in their writing tasks. As for their opinions on the portfolio assessment process, majority of the participants stated that portfolio assessment is a fair method, and is more realistic than traditional assessment tools. In addition, the participants believed that keeping a portfolio made them more responsible learners and increased their

motivation. On the other hand, some students stated that portfolio assessment increased their workload and made them exhausted since they had to complete various tasks in three skills within a limited time. Finally, they claimed that performing self-assessments and reviewing tasks were challenging for them.

In another study conducted in Iran, Nezakatgoo (2011) investigated the effects of portfolio assessment on a group of EFL students' writing performance. The study specifically focused on how portfolio-keeping affected learners' accuracy in the use of mechanics in writing. In this quasi-experimental study, 40 participants were equally placed in experimental and control groups. None of the participants in the study had a portfolio-keeping experience before and they all took a pre-test to determine their existing knowledge at the beginning of the treatment process. The participants in the control group wrote essays and submitted them to the teacher for grading. Then, the teacher assigned grades for these essays and returned them to the students. In addition, the teacher provided some oral feedback for the errors. On the other hand, the participants in the experimental group were introduced portfolio assessment method at the beginning of the term. According to this method, the participants submitted their essays but were not given a grade immediately. Instead, they were provided feedback by the teacher and revised their writings accordingly. During the feedback sessions, the teacher directed students' attention on the mechanics of writing such as punctuation, spelling, grammar, capitalization, and use of numbers. Since the participants kept their works in their portfolios, they had ample time to reflect on their errors and learn from the feedback provided on their papers. Finally, the participants selected their best works and submitted them to the teacher for grading. They were also credited for the revisions that they made. The study lasted 16 weeks and the participants in both groups took a post-test at the end of the treatment process. The statistical analyses revealed that participants in the portfolio group made significantly fewer errors in the use of mechanics in their writing. The researchers claimed that the use of portfolios helped the participants to improve their writings since they had a chance to reflect on their strengths and weaknesses throughout the process. In addition, the participants were observed to take on responsibility for their learning and became more autonomous learners.

Moradan and Hedayati (2012), in a similar vein, conducted a study to investigate the impact of two alternative assessment methods, namely portfolios and conferencing, on Iranian EFL learners' writing skills. The participants consisted of 92 intermediate EFL students who were majoring in different subjects at university. At the beginning of the study, these participants were assigned to two experimental groups and one control group. Since it

was an experimental study, a pre-test in the form of a topic-based paragraph writing was applied to three groups prior to the treatment process. The participants of the first experimental group were required to keep their written products (i.e., four types of paragraphs) in their portfolios. In addition, they were supposed to assess themselves through a self-assessment checklist. As for the participants in the second experimental group, they were supposed to participate in four whole class conferences together with two individual conferences with the teacher after completing each paragraph type. On the other hand, the teacher adopted a traditional approach in the control group and merely scored the participants' papers without providing any detailed written or oral feedback. The results of the study revealed that there was a significant difference between the performance of the two experimental groups and that of the control group on the posttest. However, no significant difference was found between the performance of the two experimental groups after implementing portfolios and conferencing techniques. In this study, the researchers emphasized the importance of keeping a portfolio in writing courses. They also claimed that students could receive feedback both from their classmates and the teacher through conferences to improve their writings.

In the Turkish context, Özer and Tanriseven (2016) conducted a small-scale study to investigate the effects of portfolio assessment on a group of Turkish EFL learners' writing achievement levels and writing self-efficacy. The participants of the study included 32 students with pre-intermediate levels of English, and they were all preparatory class students at a state university in Türkiye. The study adopted a mixed-method research design and the treatment process lasted 14 weeks. Prior to the treatment process, the participants were given a self-efficacy scale to determine their existing levels of self-efficacy. Then, the participants were provided information about portfolio assessment and its key components such as corrective feedback, redrafting, and grading rubrics. During the treatment process, the participants wrote two draft paragraphs, and they received written corrective feedback for each of their drafts. They were supposed to keep all their drafts in their portfolios, so they had the chance to reflect on their strengths and weaknesses in these drafts by looking at the comments and feedback provided by the teacher. As for grading, the first draft of each assignment constituted 30% of the total grade while the second draft made up the remaining 70%. At the end of the process, the same self-efficacy scale was administered to the participants. In addition, focus-group interviews were conducted with the participants. At the end of the statistical analyses, it was found that there were significant differences between the assignment scores of the participants. In other words, the writing performances

of the participants got better as the treatment process continued. As for their writing self-efficacy, the paired-samples t-test analyses revealed that there was not a significant difference between the pre-test and post-test scores of the participants based on the scale data. However, the focus-group interview data suggested that the participants felt more confident in their writing classes as a result of the portfolio assessment process.

In her thesis, Pesen (2016) attempted to find out the impact of Drama and Music Portfolio (DMP) on speaking skills of a group of Turkish EFL learners. Designed as a case study, the study included a single experimental group with 34 9th grade students. Prior to the treatment process, a questionnaire was administered to the participants to find out their attitudes and perceptions towards the advantages of DMP as an alternative assessment method. In addition, a pretest was given to the participants to determine their existing oral proficiency before the implementation started. At the beginning of the treatment process, the researcher introduced the basic components of music and drama portfolio in classroom since the participants were not familiar with either of these applications. For the music portfolio, the researcher brought some sample music portfolios to the classroom so that the participants could have an idea about what they were expected to do in the process. Then, the participants were divided into groups, and were asked to select a song that was compatible with the objectives of the lesson. They had several rehearsals and the teacher provided guidance and support during these rehearsals. The participants had the chance to work on their pronunciation, learned some new vocabulary and got familiar with new grammar structures during this process. Some of the participants preferred demonstrating their performances in classroom while others submitted a recorded versions of their songs to the researcher. The participants were asked to write a reflection paper regarding the whole process, and they also included the song lyrics, some rehearsal photos, and other materials in their music portfolios. Their performances and portfolios were assessed by the researcher through a grading rubric which was introduced to the participants at the beginning of the treatment process. For the drama portfolio, the participants experienced a similar process in terms of the preparation stage. The participants were asked to work in groups and wrote some scripts considering the objectives of the lesson. They had some rehearsal sessions and the researcher provided guidance and support during these rehearsals. As in music portfolio, the participants wrote a reflection paper regarding the process and included their scripts, photos, and other materials in their drama portfolio. While some of the participants recorded their performances in the classroom, the others preferred to demonstrate their performances in front of their friends and teacher. The researcher, also the teacher of the participants, graded these performances

using a grading rubric. After the treatment process, a post questionnaire was applied to the participants so that they could evaluate the process of the DMP implementation. By doing so, the researcher aimed to find out the effects of DMP on the participants' cognitive, affective, behavioral, and most importantly their speaking skills. In addition, a posttest was given to the participants to determine their oral proficiency levels after the implementation of DMP. Finally, the researcher distributed a feedback sheet to the participants to collect their views and feelings regarding the implementation of DMP and the speaking exams. The quantitative analyses revealed that the participants developed positive attitudes towards the use of DMP as an alternative assessment method in terms of improving their speaking skills. Indeed, the comparison of pre- and posttest results indicated that there was a significant increase in the participants' speaking scores after the implementation of DMP. The qualitative analyses also revealed that a great majority of the participants held positive attitudes towards the use of DMP and would like to be assessed through such methods rather than traditional speaking exams. Based on these findings, the researcher believed that teachers could utilize DMP or other alternative assessment methods to improve their students' personal and speaking skills in EFL classrooms.

In another study, Kalra, Sundrarajun, and Komintarachat (2017) investigated the effects of portfolios on the development of a group of EFL learners' writing skills. The participants of this study included 56 Thai undergraduate students who were in their senior year at an International University in Thailand. The study adopted a quasi-experimental research design with a pre-test and post-test, and the participants were assigned to a control and an experimental group equally. Prior to the treatment process, the participants in the experimental group were informed about the purpose of portfolio keeping. During the writing lessons, the participants in both groups were expected to write different types of essays such as cause-effect, compare-contrast, classification, and argumentative essays. In the control group, the teacher only taught the structure of the essay types and assigned writing tasks. However, the students were not provided any detailed feedback to revise their writings. On the other hand, the participants in the experimental group were provided feedback from the teacher and revised their paragraphs regularly. As a result, they had several drafts in their portfolios. To see the difference within groups, a paired samples t-test was conducted, and the results revealed that there was a statistically significant difference between the pre-test and post-test scores of the participants in experimental group while no such significant difference was observed in the pre-test and post-test scores of the participants in the control group. In addition, independent samples t-test results indicated

that the post-test scores of the experimental group were significantly higher than those of the control group. Based on these findings, the researchers concluded that portfolio keeping had a positive effect on the writing performance of the participants since they had an opportunity to revise their writings again and again thanks to the feedback they received from their teachers.

A recent study by Farahian and Avarzamani (2018) aimed to find out the impact of portfolio on EFL learners' metacognition and writing performance. The participants of the study were 69 undergraduate TEFL (teaching English as a foreign language) students who were studying at a university in Iran. These participants were randomly assigned to the control and experimental group. At the beginning of the study, the participants were given a metacognitive writing questionnaire, which included three sections: Person, Task, and Strategic Knowledge. In addition, they were given an essay writing task. During the treatment process, both groups were given the same amount of writing tasks, which included different essay types. The topics of the essays were selected based on students' needs and preferences. The participants in the control group wrote essays and received feedback from the teacher in terms of both the content and form of their papers. However, they did not have to keep a portfolio as a part of the course requirement. On the other hand, the participants in the experimental group were asked to make the necessary revisions to their first drafts based on the feedback they received from the teacher. Furthermore, they were given a reflection sheet to share their opinions and feelings regarding the treatment process. Since they had to keep their first and second drafts in their portfolios along with their reflection sheets, they had ample time to monitor their progress and work on their weaknesses. The results of the study revealed that there was a significant difference in the mean scores of the participants in both groups, with experimental group participants achieving higher scores. In terms of the metacognitive factors, a significant difference was observed between the two groups in terms of Person and Strategic Knowledge while no difference was observed in Task Knowledge. Finally, the participants in the experimental group developed a positive attitude towards portfolio implementation.

In a more recent study, Fathi, Derakhshan, and Safdari (2020) attempted to find out the effect of portfolio assessment on writing performance and writing anxiety of a group of Iranian EFL students. To achieve this aim, the researchers recruited 41 EFL learners as the participants of the study. The experimental group consisted of 21 participants while the control group included 20 participants. Before the treatment process, all of the participants took a pre-test, which included an essay writing task. Additionally, they completed a Second

Language Writing Anxiety Inventory (SLWAI). In the experimental group, the teacher adopted a portfolio-based writing instruction in which the participants were expected to complete timed-writing tasks and keep them in a portfolio. In addition, they received continuous feedback from the teacher and composed several drafts before submitting their best works. On the other hand, the participants in the control group received the usual writing instruction. In other words, these participants were also assigned the same timed-writing tasks and provided feedback by the teacher. However, they were not expected to produce several drafts and keep them in a portfolio. The researchers conducted ANCOVA analysis and found that keeping a portfolio in the writing lesson helped the participants improve their writing performance to a great extent. However, no such improvement was observed in the control group. In addition, the anxiety levels of the participants were significantly reduced through the use of portfolio assessment while the traditional writing instruction did not have such an effect on L2 writing anxiety of the participants in the control group. The researchers of the study concluded that adopting a portfolio-based writing instruction was beneficial for the participants since it enabled them to detect their strengths and weaknesses during the process. In addition, portfolios also provided valuable information about the individual needs of the participants. Therefore, the teacher was able to provide personalized comments and feedback for the participants' written works. Finally, the researchers claimed that portfolio-based writing instruction was effective in lowering the writing anxiety of the participants and enabled them to develop positive attitudes towards the writing lessons.

All these studies have indicated that adopting portfolio as an alternative assessment tool in language classrooms can yield positive results for learners. While commonly employed in writing classes, portfolios also enable learners to experience a systematic and organized learning process in four language skills. Portfolios also enable learners to receive continuous feedback from the teacher and improve their work in a specific skill. In addition, learners do not suffer from time pressure or exam anxiety during portfolio activities because they receive support and guidance from both their peers and the teacher. In this way, they can reach their utmost potential by the end of the process. Therefore, it is suggested that teachers make use of portfolios in language classrooms to ensure an effective teaching and learning atmosphere. With this idea in mind, the current study employs portfolio as an alternative assessment tool in writing classes following a process-based approach.

2.8.1.2. Self-Assessment and peer assessment studies. Several studies have been conducted to examine the effects of self-assessment and peer assessment on EFL learners'

academic achievement. In one of the earliest studies, Javaherbakhsh (2010) attempted to investigate whether self-assessment technique has an impact on the development of Iranian EFL learners' writing skill. The participants of the present study were 76 students with advanced level of English in two language institutions in Tehran. There were 40 participants in the experimental group whereas the control group included 36 participants. Prior to the treatment process, a writing task was given to both groups as a pre-test. Throughout the treatment process, the participants in both groups were expected to write essays containing approximately 150 words. The participants in the experimental group assessed their papers through a checklist and identified their strengths and weaknesses. Then, the teacher provided written corrective feedback on their papers as well. On the other hand, the participants in the control group only received feedback from the teacher. As the last step, the participants in both groups were given an essay writing task as a post-test. The findings of the study revealed that there was a significant difference between the post-test scores of the participants in both groups. In other words, the use of self-assessment had a positive effect on the participants' performance. Therefore, the researchers concluded that self-assessment should be implemented in EFL writing classes since it provides an opportunity for students to reflect on their writing performances.

In their small-scale study, Baleghizadeh and Zarghami (2012) investigated the impact of conferencing on a group of Iranian EFL learners' grammar learning. Conferencing was used as an alternative assessment method in the study. The participants of the study were 42 intermediate level university students who were randomly placed in experimental and control groups. Prior to the study, a 50-item grammar test and an attitude scale regarding grammar learning were applied to the participants in both groups as a pre-test. Throughout the treatment process, the participants in the experimental group had four individual conferences with the teacher. During these conferences, the teacher used a conference checklist and provided feedback for learners' strengths and weaknesses in several grammar issues. In addition, the participants had four whole-in-class conferences where they exchanged ideas and comments with their peers through the same conference checklist on related grammar issues. At the end of the treatment process, the same pretests were given to the participants in both groups. The statistical analyses revealed that there was a significant difference between the posttest scores of the participants in both groups, with experimental group participants achieving higher scores. In addition, a significant difference was observed between the pre and posttests scores of the participants in the experimental group based on

the scale data in grammar learning. In other words, the participants in the experimental group developed a positive attitude towards learning grammar.

In her thesis, Medfouni (2014) examined the effect of self and peer assessment on a group of EFL students' essay writing skills. Furthermore, the author tried to reveal the accuracy level of these assessments compared to teacher assessment. The study had a quasi-experimental design, and the participants were 40 EFL students who were studying at a university in Algeria. Both the experimental and control group included 20 students with similar characteristics and writing background. At the beginning of the study, a pre-test in the form of an essay writing task was given to the participants in both groups. The participants in the experimental group were assessed through *the four-by-four method*, which was proposed by Anderson and Bachor (1998). Following this method, the teacher and students had a discussion on the rubric for the assessment of comparison and contrast essay. After the final version of the rubric was obtained, the participants were then asked to write their first essays in the classroom. Next, they were asked to assess their papers using a checklist which was prepared in parallel with the rubric. The participants were then asked to form groups of four to assess each other's essays based on this checklist. In other words, they performed peer assessment and provided feedback for their peers. In the following days, the participants wrote two more essays and performed self-and peer assessment in the same manner. The essays written by the participants were also assessed by the teacher to reveal the accuracy of the participants' assessment, which was an important aim of the study. As for the control group participants, they were also engaged in the same essay writing activities. However, the teacher only provided some written feedback on these essays in a traditional manner and returned them to the participants afterwards. After the treatment process, a similar essay writing task was given to the participants in both groups as a post-test. The results of the study indicated that the participants in the experimental group achieved significantly higher post-test scores than the participants in the control group. In addition, great similarities were observed between the teacher assessment and the participants' assessment in the experimental group. Based on these findings, the researcher concluded that teachers could utilize self and peer assessment in EFL classes to improve students' writing performances. It is also suggested that if teachers make use of self- and peer assessment, they could promote interaction and collaboration among students in their classrooms.

Iraji, Enayat, and Momeni (2016), in a similar vein, conducted a study to investigate the effect of self- and peer assessments on a group of Iranian EFL learners' argumentative essay writing performance. The participants of the study included 36 intermediate level of English learners who were studying at a higher education institution in Iran. Having been randomly assigned to experimental and control groups, the participants were asked to write a five-paragraph argumentative essay as a pre-test. Throughout the treatment process, the participants in the experimental group were engaged in both self- and peer assessments after they wrote essays. They used a checklist and a rubric during these assessments. In addition, the teacher also provided feedback and comments on their papers. On the other hand, the participants in the control group merely received feedback from their teachers, which was the usual practice in the institution. After the treatment process, the participants in both groups were given another argumentative essay writing task as a post-test. The independent t-test analyses revealed that the participants in the experimental group achieved significantly higher scores than the participants in the control group. Likewise, the paired samples t-test results indicated that there was a significant difference between the pre- and post-test of the participants in the experimental group while no such significant difference was observed in the scores of the participants in the control group. The researchers concluded that self- and peer assessments had a positive effect on the argumentative writing performance of the participants and suggested that teachers need to provide more opportunities for self- and peer assessments in the writing lessons.

In a more recent study, Ghazizadeh and Bazargani (2019) attempted to find out the effect of self- and peer assessments on the writing ability of a group of Iranian EFL learners. The study had a true experimental design and the participants of the study consisted of 45 female students who were randomly assigned to two experimental groups and one control group. At the beginning of the study, a paragraph writing task was given to all of the participants as a pre-test. During the treatment process, the participants in the first experience group were engaged in self-assessment while the participants in the second experimental group utilized peer assessment. During the assessment process, the participants in both groups used an analytical rubric, which consisted of five sections: content, organization, language in use, grammar, and mechanics. On the other hand, the participants in the control group merely received feedback from the teacher in a traditional manner. In other words, the teacher wrote some comments on their papers and provided some oral feedback as well. The treatment process lasted 10 sessions and post-test was given to all participants at the end of the study. The results of the study revealed a significant difference between self, peer, and

traditional teacher assessment groups. In addition, it was found that the participants in the peer assessment group achieved significantly higher scores than the other participants in two groups. Finally, the participants in the self-assessment group were observed to outperform the participants in the traditional teacher assessment group. Considering these findings, the researchers of the study strongly believed that self- and peer assessment had a positive effect on the paragraph writing abilities of the participants. Therefore, students need to be more actively involved in the assessment process in EFL classrooms.

In the Turkish context, Kızıllı (2019) conducted a seven-week case study in which he integrated peer assessment into EFL writing classes and aimed to find out whether this type of assessment enabled learners to improve their social skills. The participants of the study were 17 students who were studying at an English preparatory program of a foundation university. Prior to the study, the participants were introduced to a grading rubric to assess a written work. During the first three weeks of the treatment process, the participants wrote different types of paragraphs and assessed their peers' writings based on the rubric. Then, they wrote different types of essays for the remaining four weeks and followed the same assessment procedure. Apart from grading their peers' papers based on the rubric, they also completed a peer assessment form. Throughout the process, the participants were interviewed by the researcher twice at different times so that their opinions and feelings regarding the assessment procedures could be gathered. The qualitative analyses revealed that students found peer assessment activities very useful, and that they learned how to use the rubrics to assess their peers' papers. In addition, they enjoyed the cooperative learning atmosphere in the classroom and benefitted from receiving feedback from each other. As a result, their collaboration and problem-solving skills were enhanced, which improved their social skills as well.

In another recent study, Fathi et al. (2021) attempted to investigate the impact of self- and peer-assessment on the writing performance of a group of Iranian EFL learners. In addition, the study tried to find out whether self- and peer assessment had any effect on the self-efficacy of these learners in writing classes. The participants of the study included 36 intermediate EFL who were randomly assigned to a self-assessment group and a peer-assessment group. There were 17 participants in the self-assessment group while the peer-assessment group had 19 participants. Prior to the study, a timed-writing essay, and a Writing Self-Efficacy Scale (WSES) were given to participants as pre-tests. Throughout the treatment process, the participants in the self-assessment group were provided information and guidance on how to assess their papers by themselves while the participants in the peer-

assessment group were taught to assess the writings of their peers. The treatment process lasted 15 sessions. After the treatment process, the participants in both groups were given another timed-writing essay, and the same self-efficacy scale as post-tests were administered to both groups. The paired- samples t-tests and ANCOVA analyses revealed that both self- and peer-assessment activities significantly improved the writing performance and writing self-efficacy of the participants. In addition, it was found that the participants in the peer-assessment group achieved significantly higher scores than the participants in the self-assessment group in terms of both writing performance and writing self-efficacy. Based on these findings, the researchers suggested peer-assessment had a more beneficial effect than self-assessment in terms of improving the writing abilities of the participants. In addition, peer assessment was more effective in boosting the self-efficacy of the participants in writing classes.

All the aforementioned studies have concluded that learners can experience a more conscious learning process through the use of self-assessment and become more aware of their strengths and weaknesses in specific language skills. They can also benefit from a cooperative learning atmosphere in the classroom by means of peer-assessment and receive feedback and support from each other. As a result, their problem-solving skills and social skills can also improve. Therefore, teachers are advised to make use of self- and peer assessment in language classrooms. With this suggestion in mind, the current study employed peer- and self-assessment as an alternative assessment method in writing classes.

2.8.2. Studies Conducted on Students' Perceptions of Alternative Assessment

In one of the earliest studies, Burnaz (2011) investigated the perceptions of a group of EFL learners towards portfolios as a method of alternative assessment. In her study, she specifically aimed to reveal the effects of portfolio assessment on learner autonomy. The study was conducted at a state university in Türkiye and the participants included 21 students with intermediate level of English. The study employed a mixed method research design. Prior to the study, a pre-survey was distributed to the participants to gather their opinions regarding their views on learner autonomy. In addition, an open-ended question survey was applied to collect their existing views on portfolio assessment. During the treatment process, which lasted 13 weeks, the participants were engaged in several activities in different skills and kept their works in their portfolios. In addition, they had to fill in a reflection paper to express their views and feelings on the treatment process. The teacher graded the participants' works based on a predetermined criteria and provided feedback for them. At

the end of the study, a post-survey was given to the participants to find out whether their opinions on learner autonomy changed. Furthermore, a structured interview was conducted with each of the participants to gather their views and feelings on the implementation of portfolio assessment. The results of the study indicated that the participants held positive attitudes towards portfolio assessment before and after the treatment process. In addition, they claimed that they became more autonomous learners after the treatment process. Finally, the participants mentioned several disadvantages of traditional assessment methods. They claimed that such methods put time pressure on them and only led to memorization and temporary learning. In addition, they believed that they could not improve their speaking skills when they were assessed through traditional assessment methods. The researcher of the study therefore suggested that teachers need to employ portfolio assessment to improve learners' four skills in EFL classrooms. In addition, she claimed that learner autonomy could be enhanced greatly through the use of portfolios.

In another study, Afshar and Bastami (2012) attempted to find out whether portfolio assessment would enhance a group of EFL learners' autonomy in language learning. 60 students, who were learning English at a private language institute in Iran, participated in the study. These participants were equally assigned to experimental and control groups at the beginning of the study. A learner autonomy questionnaire was applied to all participants prior to the study. During the treatment process, the participants in the experimental group completed written and recorded tasks and kept them in their portfolios. These participants constantly received feedback from the teacher and tried to revise their works accordingly. In this way, they had a chance to reflect on their strengths and weaknesses and assess their products by themselves. By doing so, they were expected to perform better in the following tasks. On the other hand, the participants in the control group had to complete the same tasks but did not have to keep them in a portfolio. They submitted their works and received some feedback and comments from the teacher, which was the usual practice at the institute. When the treatment process ended, the same learner autonomy questionnaire was applied to the participants in both groups. In addition, a semi-structured interview was conducted with 10 of the participants from each group. As for the results of the study, the t-test analyses indicated a significant difference in terms of participants' autonomy levels. In other words, the participants in the experimental group enhanced their autonomy more than the control group. The qualitative analyses revealed that the participants in the experimental group favored portfolio assessment since it created an opportunity for self-assessment, taking responsibility, goal setting and raising awareness. Based on these findings, the researchers

of the study suggested that portfolio assessment could provide a supportive learning environment and eliminate the pressure of one-shot exams in classrooms. In addition, it might raise learners' metacognitive awareness and help them discover their best ways of learning. Finally, learners could take on responsibility for their own learning and assess their own performances throughout the learning process.

In her thesis, Cirit (2014) investigated the perceptions of pre-service EFL teachers on the integration of web 2.0 tools to their courses as an alternative assessment method. In addition, the study aimed to reveal the perceptions of these pre-service teachers regarding traditional, online, and alternative assessment methods. The participants of the study consisted of 40 second grade pre-service ELT teachers in Türkiye. Prior to the study, a pre-survey was given to the participants to reveal their attitudes toward technology and three types of assessment (i.e., traditional, online, and alternative assessment). Throughout the treatment process, the participants had to complete six tasks that were designed by the researcher. These tasks were prepared based on the syllabus of "ELT Methods I" course. The participants were supposed to utilize different types of Web 2.0 tools such as Voki, Testmoz, Mindomo, Glogster, Facebook and Prezi to complete these tasks. After each task, the participants were supposed to fill in a reflection paper to express their views and feeling regarding the tasks. As for the feedback sessions, the researcher utilized different types of feedback for each task. For some tasks, self-, pair or whole class assessments were used to provide feedback. On the other hand, computer-based feedback sessions were also held for some of the tasks. The researcher designed a different rubric to evaluate each task and the grades the participants got from these tasks constituted 30% of their overall grades in the course. After the treatment process, a post-survey was applied to the participants to see whether their attitudes towards the three types of assessment have changed. The post-survey also included some extra sections to gather the participants' views and feelings about the treatment process including the tasks and feedback sessions. As the last step, a semi-structured interview was conducted with the four participants who completed all the requirements of the study. The purpose of the interview was to gather an in-depth data on the three assessment methods and the treatment process. The results of the study revealed that the participants held positive attitudes towards the use of alternative assessment through Web 2.0 tools prior to the study and their attitudes got even more positive after the treatment process. As for the three assessment types, the participants favored alternative assessment more than online or traditional assessment. They believed that their motivation in the methodology course was increased by the use of Web 2.0 tools as an alternative assessment

method. They also stated that alternative assessment enhanced their learning and promotes critical thinking skill since it provided continuous detailed feedback for them to reflect upon.

In another thesis, Fajarsari (2016) attempted to examine the perceptions of a group of Indonesian EFL learners toward certain alternative assessment methods. In addition, the study aimed to find out these EFL learners' favorite alternative assessment types. The participants of the study included a total of 40 11th and 12th grade students. The researcher selected these participants since they were already familiar with the use of different alternative assessment methods in their schools. To collect data, a questionnaire was given to the participants to collect their views and feelings regarding the use of certain alternative methods. Specifically, the questionnaire included items concerning the use of self-assessment, peer assessment and performance assessment as alternative assessment methods. The statistical analyses revealed that the majority of the participants held positive attitudes towards the use of alternative assessment methods along with traditional ones in EFL classes. The participants believed that alternative assessment methods foster the use of four language skills in an integrated way. In addition, they believed that alternative assessment methods created a cooperative learning environment and increased their motivation to learn the target language. As for their favorite alternative assessment method, the participants favored self-assessment more than peer assessment or performance assessment though all of these assessment methods provided several benefits for them. They stated that self-assessment provided them an opportunity to check their products based on the predetermined criteria before submitting them for grading. In this way, they were able to get higher grades from the teacher. In addition, performance assessment enabled them to use the language in productive tasks and improved their productive skills. Finally, they believed that peer assessment created a cooperative learning environment in the classroom and also reduced the stress of being assessed by the teacher.

Safari and Koosha (2016) tried to investigate the effects of speaking portfolio, which is commonly used as an alternative assessment method, on a group of Iranian EFL learners' speaking ability. The study also aimed to explore the perceptions of these learners' regarding the use of speaking portfolio in their speaking classes. The study had a mixed-method research design, and the participants of the study were 64 intermediate and advanced learners who were randomly assigned to four groups. There were two experimental groups (i.e., intermediate, and advanced groups) and two control groups (i.e., intermediate, and advanced groups). Prior to the treatment process, all the participants were given a pretest to assess their existing oral proficiency levels. Later, the participants in the experimental groups were

introduced to basic components of a speaking portfolio. They were also informed about the functions and aims of using a speaking portfolio. In addition, the teacher introduced a grading checklist and a reflection paper which were used to assess participants' performances. The treatment process lasted 12 sessions and for each session, the participants in the experimental group were assigned individual, pair or group speaking tasks. These tasks were designed by the teacher considering the topics in the coursebook as well as the objectives of the lesson. During the sessions, the teacher video recorded the performances of the participants and displayed these recordings to the participants so that they could assess their performances using the grading checklist. In addition, the participants were asked to provide feedback on their peers' performances. Finally, the teacher graded the participants' performances and their reflection papers to determine their final scores. All these performances, reflection papers and assessments were stored in the participants' speaking portfolios. To increase the reliability of the scores, another rater also graded the participants' performances using the same grading checklist. On the other hand, the participants in the control groups were assigned the same speaking tasks, but the teacher did not record their performances to create an opportunity for self- or peer assessment. The teacher graded these participants' performances using the same grading rubric. At the end of the study, all the participants were given a posttest to compare their oral proficiency levels before and after the treatment process. The reliability of the pre and post test scores were ensured with the help of another rater. Finally, a questionnaire was administered to the participants in the experimental group to find out their attitudes towards the use speaking portfolio as an alternative assessment method in their speaking classes. The statistical analyses revealed that both intermediate and advanced students in the experimental groups achieved significantly higher scores than their counterparts in the control groups. Finally, the participants in the experimental groups were observed to develop positive attitudes towards the use of speaking portfolios since it created a space for self-reflection and the participants were able to detect their strengths and weaknesses during self- and peer assessment sessions. The participants also claimed that they improved their speaking skills dramatically and felt more confident in speaking lessons as the time passed.

In another study, Ardianti and Mauludin (2017) attempted to examine the students' responses in an EFL reading class where authentic assessment was implemented to help them improve their reading skills in English. The participants of the study were 27 EFL learners studying at a university in Indonesia. To achieve the purpose of the study, a qualitative case study was applied, and data were collected from questionnaire and 16

meeting field-observation. During the treatment process, the participants were assigned several reading tasks, which required them to work in pairs or groups. During these activities, they were asked to assess their own performances and provide comments for their peers' performances as well. They had several group discussions and made a number of presentations based on the texts assigned. In this way, they were able to use the language in an integrated way. The qualitative analyses revealed that the students gradually responded positively to the application of authentic assessment. It was found that authentic assessment motivated the participants to have a reading routine, which enabled them to expand their vocabulary and get familiar with English discourse. In addition, the participants had a chance to demonstrate their reading comprehension through productive tasks such as writing reviews, presentations, and discussions. Finally, authentic assessment enabled them to monitor and assess their own reading performance. Based on these findings, the researchers concluded that most of the participants enjoyed the implementation of authentic assessment in reading classes through a series of social practices.

In Turkish context, Cetinkaya (2017) examined Turkish EFL learners' perspectives regarding the use of portfolio assessment in their writing courses. The participants of this descriptive study included 50 preparatory class students who were expected to major in English Language and Literature Department in the following year. The study adopted a mixed method research design and employed a questionnaire and semi-structured interviews to collect data. Throughout the treatment process, the participants had a 4-hour writing class each week. During the first term, they learned how to write different types of paragraphs. These included narration, process, classification, compare/contrast, and opinion paragraphs. On the other hand, they focused on how to write opinion and argumentative essays. The participants received feedback from the teacher and revised their written works if necessary. They kept all of their drafts in their portfolios, and these were assessed by the teacher based on a predetermined grading rubric. The quantitative analyses of the study indicated that the participants developed positive attitudes towards portfolio assessment for several reasons. First of all, they had an idea of their own language improvement and were able to detect their strengths and weaknesses in writing. In addition, portfolio assessment increased their sense of responsibility and promoted learning outside of school as well. On the other hand, some of the participants complained that their workload increased due to portfolio assessment and thus they opted for traditional assessment instead. In parallel with the findings of the quantitative analyses, the qualitative analyses also reported several benefits of using portfolio assessment in EFL classes along with some drawbacks. As for the positive aspects,

the participants stated that portfolio assessment reduced their anxiety levels in writing classes since it created a dynamic and cooperative learning environment. In addition, they claimed that portfolio keeping increased their motivation in writing and encouraged them to write more both in and outside class. On the other hand, some of the participants complained about the amount of work that they had to do as a part of portfolio assessment. In short, the researcher of the study concluded that portfolio assessment mostly affected learners' attitudes towards writing in a positive way. However, she also mentioned that implementing portfolio assessment in writing classes could create some challenges for students.

In her MA thesis, Özuslu (2018) attempted to investigate a group of EFL learners' perceptions regarding the use of Performance-based Assessment Tasks (PTs) as a supplementary component of the assessment system in a preparatory school in Türkiye. In addition, the study examined the expectations and views of the instructors and administrators on the implementation of PTs. The study had a mixed method research design, and the participants consisted of 126 students, 60 instructors and two administrators. Prior to the treatment process, an open-ended questionnaire was given to the participants (i.e., the students, instructors, and administrators) to collect data on the expectations of all these stakeholders concerning performance-based assessment prior to the implementation of PTs. Through this questionnaire, the participants had the chance to express the possible benefits and challenges of PTs. The whole treatment process lasted 15 weeks. The first three weeks were spent on the orientation and training of all participants regarding the implementation PTs. In the following weeks, the participating students completed several PTs for their grammar, reading/vocabulary, and listening/speaking lessons. Due to the nature of performance assessment, both analytical and holistic rubrics were used to assess the performance of the participating students. The qualitative data gathered from the participants were analyzed separately through content and thematic analysis while the data collected through the scales was analyzed by means of a statistical program. The open-ended items in the survey were also analyzed through content and thematic analysis method. The analyses of the qualitative data collected prior to the treatment revealed that all the participants mentioned the potential benefits of PTs for language learning. On the other hand, the participating students stated some challenges such as time limitations, amount of effort and difficulty of PTs. Similarly, the participating instructors indicated assessment-related concerns while the participating administrators mentioned the excessive workload of the instructors. The data collected after the treatment process revealed that the all the participants had a positive attitude towards the use of PTs in terms of planning, application, scoring,

learning outcomes and program consistency phases. Finally, although majority of the stakeholders recommended some revisions for the PTs, they all agree that they should be a part of the assessment system in the preparatory school in the following years. Based on these findings, the researcher of the study believed that PTs need to be considered as a supplement to traditional assessment methods at preparatory schools. However, she also suggested that a further study can be conducted to investigate the impact of PTs on the traditional exam scores of students to see whether PTs have any impact on students' academic achievement.

In a more recent study, Ghoorchaei and Tavakoli (2020) aimed to explore the impact of portfolio assessment on a group of Iranian EFL learners' L2 writing and its subskills such as focus, elaboration, conventions, organization, and vocabulary. In addition, the study aimed to reveal these EFL learners' perceptions regarding the use of portfolio assessment in writing classes. The study had a mixed method research design, and the participants of the study comprised 30 university students who were majoring in English Literature and Translation departments. During the treatment process, the participants learned key issues in essay writing such as how to write a thesis statement, body paragraphs and conclusion etc. Having learned these basic concepts in essay writing, the participants wrote five different types of essays during the treatment process. After each essay type, they assessed their own papers using a scoring rubric. Then, they submitted their papers and received feedback from the teacher. In addition, they assessed their peers' papers using a peer review guide. In short, the portfolio assessment procedure included self-, peer and teacher assessment. The participants were asked to revise their essays based on the feedback and comments provided by their peers and the teacher. To determine their final grades, the teacher asked the participants to select three out of five of their final drafts in their portfolios. The participants' views and feelings regarding the implementation of portfolio assessment were gathered through a "portfolio contribution questionnaire and semi-structured interviews. The quantitative data analyses revealed that most of the participants have improved their writings, especially in terms of focus, elaboration, and organization. The qualitative data analyses indicated that the participants hold positive attitudes toward portfolio assessment in writing since it provided several benefits for them. They claimed that they not only enjoyed the product but also the process of writing thanks to the use of portfolio assessment. Therefore, they felt more motivated in writing classes and enjoyed working in a collaborative learning environment.

In her MA thesis, Piri (2022) attempted to explore and compare the perceptions of a group of Turkish EFL teachers and students on the use of electronic portfolio (EP) as an alternative assessment method in writing lessons. The participants of the study included 30 EFL students and 31 EFL teachers who worked at a preparatory school of a foundation university in Türkiye. The study had a qualitative research design, and the data were gathered by means of open-ended questionnaires and focus group interviews. The participants of the study completed several writing tasks and submitted them to the teacher through different online platforms. All of these tasks were checked for plagiarism through a Turnitin tool, and the participating teachers provided feedback for these tasks. The participating students then revised their tasks and uploaded them to the online platform again. By the end of the treatment process, all of the participating students had an EP which can be accessed by both teachers and students anytime anywhere. The qualitative analyses revealed that most of the participating teachers and students viewed EP as a beneficial alternative assessment tool in their writing classes. The participating students stated that EP was effective in improving their writing skills, especially in terms of organization, sentence formation, vocabulary, grammar, and appropriate use of punctuation. In addition, they believed that they felt more confident in their writing abilities. On the other hand, they believed that EP assessment could be improved especially in terms of providing feedback and topic selection. Similarly, the participating teachers stated that EP was a practical tool to store students' works and provide feedback for them. In addition, they believed that EP created an interactive learning environment for students. As one can see, most of the teachers and students who participated in the study favored EP as an alternative assessment method and suggested that EFL teachers working at preparatory schools in Türkiye ought to implement EP in their writing classes as an alternative assessment method.

All in all, a great number of studies have been conducted on the use and effectiveness of alternative assessment methods in EFL classes. Some of these studies attempted to investigate the impact of alternative assessment methods on EFL learners' academic achievement levels whereas other studies aimed to explore students' perceptions about the use of alternative assessment methods. However, there have not been many studies that investigated the impact of alternative assessment methods on students' language achievement levels and the perceptions of these students regarding the use of such methods in EFL classes in Türkiye, particularly in the school of foreign languages context. Therefore, the purpose of the study is to find out the effect of applying alternative assessment methods on a group of Turkish EFL students' academic achievement and investigate the attitudes of

these students towards learning English before and after the implementation of alternative assessment methods in a preparatory school. Based on the findings of the study, the researcher also aims at proposing an alternative assessment implementation guideline for preparatory schools in Turkey. In this sense, the present study could create awareness and pave the way for new alternative assessment practices in these schools.

CHAPTER 3: METHODOLOGY

This chapter explains the methodology of the study in five different sections. First of all, the research design is described, and then the setting and participants of the study are introduced in detail. Next, the data collection instruments are introduced, and the data collection process is described in detail. Finally, the data analysis section is presented.

3.1. Research Design

The purpose of this study is to find out the effect of applying alternative assessment tools on a group of Turkish EFL students' academic achievement and investigate the attitudes of these students towards learning English before and after the implementation of alternative assessment tools. Forming the research design was a crucial step of the study since a research design determines some key elements such as the participants, data collection instruments, treatment process, and data analysis. In this sense, it serves as the backbone of the research study by holding these key elements together in harmony (Trochim, 2005). To achieve the purpose of the study, mixed-method research design was adopted. In other words, both quantitative and qualitative methods were used to collect data during the study so that a more complete understanding of the research problems could be provided.

In a quantitative research design, the researcher attempts to explain a subject matter through numerical data which are usually gathered as a part of an empirical research study. Many researchers use this research design to test a theory that consists of several variables. In quantitative research, researchers generally gather data through surveys, questionnaires, or tests to obtain numerical data. They then analyze the numerical data through some statistical programs and see whether their theory is true or not (Cresswell, 2007). There are legitimate reasons to use a quantitative research design in social sciences. First of all, researchers are concerned with the objectivity of their findings. They seek to obtain findings without the interference of their previous experiences, biases or even world views. In addition, quantitative research aims to reach generalizable or replicable results. The validity and reliability of numerical data empower the findings of quantitative research, which leads to generalizable and replicable interpretations (Harwell, 2011). However, many scholars have criticized quantitative research design since it neglects the views, beliefs and feelings of people while conducting a study and interpreting its findings (Bryman, 2016).

Due to the criticism towards quantitative research design, some scholars have suggested using qualitative research design. In a qualitative research design, researchers

investigate an issue in its natural setting and collect direct and in-depth data. As Bogdan and Biklen (2007) claim, the context plays a key role in shaping people's attitudes or behaviors towards a specific issue. Therefore, it is of great significance to collect data in the natural setting of people. In addition, researchers tend to use multiple forms of data in a qualitative study, so they are able to have a better understanding of the issue that is investigated through the triangulation of the data (Denzin, 1978). According to Cresswell (2012), qualitative research yields meaningful results, since it includes a systematic data collection process through in-depth interviews, observations, documents, and other data sources. The researcher also has a major role in qualitative research design since the interpretation of data is handled by the researcher. Finally, the qualitative research design does not have a pre-determined research plan. Therefore, some phases of the study might be altered during the process (Cresswell, 2009). Similar to quantitative research method, the qualitative research method has certain weaknesses. First of all, the findings of a qualitative research might be questioned by readers in terms of validity and reliability since the analysis and interpretation of the data are handled by the researcher/s only. Second, the generalizability of the findings may not be possible since data is collected from a specific group in a specific context. Finally, each phase of qualitative research requires a huge amount of time and effort, so it may not be favored by some researchers.

To compensate for the weaknesses of quantitative and qualitative methods, researchers have suggested using a combination of the two methods. As a result, a new term called mixed methods has emerged in the literature. According to Cresswell (2007), mixed methods can be defined as a procedure whereby the researcher gathers, analyses, and finally combines or mixes both quantitative and qualitative data within only one study. Researchers have used mixed methods research designs since it has several strengths. First of all, it can help researchers to clarify or explain complex phenomena using both quantitative and qualitative data. In this way, they can have a detailed understanding of the research problem (Cresswell, 2012). In addition, they confirm or cross-validate the findings of different data analyses on the same issue through triangulation (Denzin, 1978; Rossman & Wilson, 1985). Finally, mixed methods research design increases the validity and reliability of the findings of a study. Therefore, it is favored and welcomed more than quantitative or qualitative research designs in social sciences.

There are different types of mixed method designs in the literature. In the present study, a sequential explanatory research design was employed. As Figure 3.1 shows, a researcher collects quantitative data for the study but then requires qualitative data to refine

or enrich the quantitative findings (Cresswell & Plano Clark, 2006). These two types of data are analyzed at different times and the results of the qualitative analysis are used by the researcher to get an in-depth understanding of the quantitative data. As the last step, the interpretation of the entire data analysis is carried out by the researcher. In this sequential mixed-method study, which adopted a quasi-experimental design due to convenience sampling, quantitative data were collected through a scale and the school tests while the qualitative data were obtained by means of semi-structured interviews and learning journals.

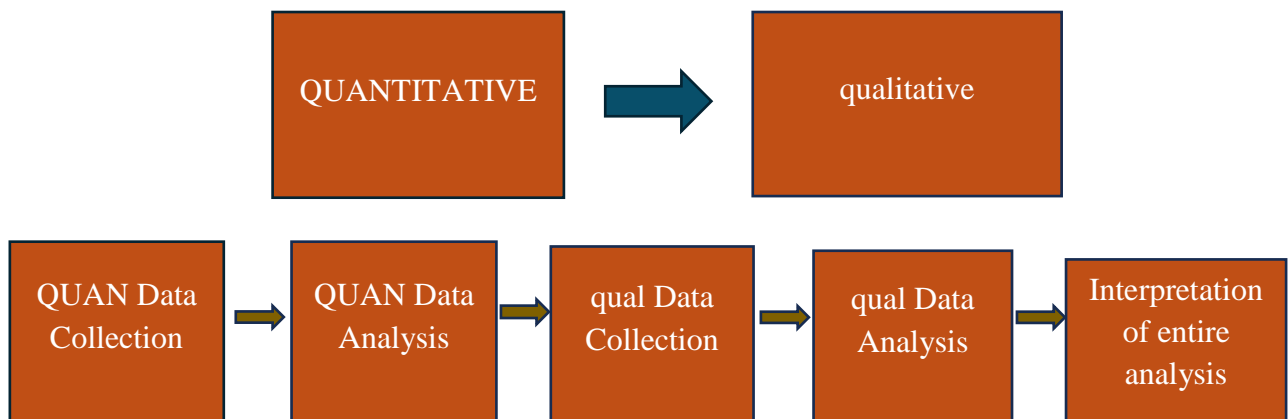


Figure 3.1. A Sequential explanatory design (Cresswell, Plano Clark, Guttman & Hanson, 2003).

3.2. Setting and Participants

The current study aims at finding out the effect of applying alternative assessment tools on a group of Turkish EFL students' academic achievement and investigating the attitudes of these students towards learning English before and after the implementation of alternative assessment tools in a preparatory school in Türkiye. The data were collected during the Spring semester of the 2021-2022 academic year in the School of Foreign Languages at Pamukkale University in Denizli. In the School of Foreign Languages, students come from many different degree programs including both 100% and 30% English Medium Instruction (EMI) students as well as the ones who take preparatory school education electively. The goal of the school is to prepare the students for their education life in their faculties by providing the necessary knowledge and skills in the target languages. In the school, the foreign language learning program is divided into four modules which are identified as A1, A2, B1, and B1+. These levels correspond to beginner, elementary, pre-intermediate and intermediate levels respectively. Each module lasts eight weeks with 24

hours of English classes per week. In a weekly schedule, five hours are allocated to reading and writing courses each while speaking and listening courses are allocated three and two hours respectively. Lastly, nine hours are allocated to core lessons, in which teachers mostly focus on grammar issues along with listening, reading and speaking skills.

At the beginning of the school year, students take a proficiency exam, which consists of one written and one speaking sessions. The weight distributions of these exams are 80% and 20%, respectively. Students whose total scores are equal to 70 and above are considered successful, and they proceed to their faculties directly. Students who fail in the proficiency exam take a placement exam, and they are placed in different levels such as A1, A2, and B1 based on their scores. As mentioned above, each module lasts eight weeks and during this period, students take one final (proficiency) exam, one midterm exam, and five quizzes, all of which are administered in a traditional pen and pencil format. To be considered successful and proceed to the next level, students' total scores need to be 70 or above. Their total scores are calculated by adding 50% of the final (proficiency) exam, 30% of the midterm exam, 10% of the quizzes and 10% of their performance worksheets. The students who are placed in the B1 level at the beginning of the term (i.e., in September) have a chance to complete their preparatory school education in January on condition that they cover the B1+ level, which is the exit level of the program. If students are placed in the A1 level, which is the case for the majority of the students, they need to complete four modules successfully to graduate from the preparatory school in June. During these four modules, they are expected to take four final (proficiency) and four midterm exams, 20 quizzes and complete nearly 40 performance worksheets. As one can see, students have a tight schedule of exams throughout the year, which could create a stressful test-oriented learning environment in the school.

The participants of this study were chosen through convenience sampling method since the researcher himself was an instructor in the School of Foreign Languages. Therefore, organizing the setting, collecting the data, and managing the treatment process was convenient for the researcher (Cresswell, 2012). The participants included 75 preparatory class students all of whom were young adult learners, and their ages ranged between 17 and 19. This quasi-experimental study had an experimental and a control group. There were 38 participants in the experimental group, while the control group had 37 participants. Both experimental and control group participants were placed in B1 classes based on the placement test administered by the School of Foreign Languages and were expected to reach B1+ level by the end of the term.

3.3. Data Collection

The researcher of this study was also the instructor of the participants in the experimental group. He has been teaching English at the college level for fourteen years. To conduct this study, certain alternative assessment tasks were designed in parallel with the learning objectives of each skill course, coursebook content, and syllabi to ensure the validity of the tasks. In addition, a number of some checklists and rubrics were selected to assess the participants' performances. Throughout the treatment process, the researcher met the participants in their reading, writing, listening, and speaking classes. He tried to implement certain alternative assessment methods and tools in each of these classes. The instructors of the control group had similar educational backgrounds and teaching experiences. Prior to the treatment process, they were provided detailed information about the purpose and implementation of the study. Most importantly, it was made sure that they would teach their lessons in their usual way without incorporating any alternative assessment methods or tools.

3.3.1. Data Collection Instruments

The present study adopted a mixed method research design and appropriate data collection instruments were determined accordingly. Prior to the study, the researcher applied for ethics committee approval on the 29th of September to collect data from the participants studying in the preparatory school. The application was approved by the ethics committee, and the approval document with an issue number as 68282350/22021/G020 was received on the 5th of November. The necessary permissions were also granted from the administration of School of Foreign Languages. To achieve the purpose of the study, both quantitative and qualitative data were collected through certain instruments. The quantitative data collection tools included an attitude scale, traditional exams, and alternative assessment tasks while the qualitative data were collected through semi-structured interviews and student learning/reflective journals.

3.3.1.1. Attitude Scale. The attitude scale (See Appendix A) was developed by Ayşegül Takkaç Tulgar, who works as an associate professor at Ataturk University, in 2018. It was a 5-point Likert-type Scale consisting of 26 items with options ranging from "Completely Agree (5)" to "Completely Disagree (1)". The scale was originally applied to 370 participants, and the analyses of internal consistency, half-split and test-retest reliability

were conducted to check the reliability of the scale. The reliability coefficient of the whole scale was .87.

3.3.1.2. Traditional exams. The tool that was used to obtain the experimental and control group participants' traditional assessment scores included a series of exams, which were prepared by the instructors working at the School of Foreign Languages. As mentioned previously, the data collection process included B1 and B1+ modules. During each module, the participants in both groups took a midterm and a proficiency (final) exam. Both midterm and proficiency (final) exams included four sections such as reading/vocabulary, writing, listening and language use, and each section was assigned 25 points. The proficiency (final) exam also included a speaking section, which constituted 20% of the participants' total final exam scores.

In the reading section, the participants were asked to read at least two texts and answer some comprehension questions about them. These questions aimed to test certain sub-skills of reading such as skimming, scanning, reading for gist, and deducing the meaning from the context in the form of multiple-choice or true/false test items. The vocabulary section also consisted of multiple-choice or gap-filling questions. In the listening section of the exam, the participants listened to two listening tracks twice and answered some multiple-choice or true false questions. Like reading section, the questions in the listening section aimed to test certain sub-skills of listening such as skimming, scanning, listening for gist, and deducing the meaning from the context in the form of multiple-choice or true/false test items. As for the writing section, the participants were asked to write a paragraph or essay on one of the topics given. The topics in the writing section were related to different genres such as compare-contrast, cause-effect, and expressing an opinion. Finally, the participants were expected to answer multiple-choice questions in the language use section of the exam. The questions in this section aimed to test the grammatical knowledge of the learners.

In addition to the midterm and proficiency (final) exams, the participants took five quizzes in each module, which tested four language skills of the participants along with their language use proficiency. The reading/vocabulary, listening, writing and language use quizzes were administered in written form in the classroom setting. Both the final speaking exam and speaking quiz were administered in the same manner. The topics and task cards that would be used in the speaking exams were shared with the participants on the website of the school. The participants studied these materials and attended the speaking exam prepared on the predetermined day. Each participant was matched with one of their peers

randomly and there were two assessors responsible for evaluating their performances in the exam. Although they took the exam with one of their peers, each participant was given a score based on their individual performances. The performance of the participants was evaluated by the two assessors based on an analytic speaking rubric provided by the Testing Office of the school.

The participants were also assigned performance assignments, which included worksheets for each skill. Although they are called performance assignments, these also included traditional question items such as multiple-choice, true-false, and gap-filling exercises. In addition, the instructors at the School of Foreign Languages were free to give performance grades based on any in-class work that their students do. However, the type of the in-class works or how it is evaluated varies enormously depending on the instructor, so there is not a standardized and systematic in-class performance assessment of the students.

As mentioned above, the participants in both experimental and control groups took a lot of traditional sit-down exams during the modules. The participants' total scores, which determined whether they successfully completed the modules, were calculated by considering the following weight distribution shown in Table 3.1.

Table 3.1. *Traditional Assessment Tools and Their Percentages*

<i>Traditional Assessment Tools</i>	<i>%</i>
Proficiency (Final) Exam	50
Midterm Exam	30
Quizzes	10
Performance Assignments	10

3.3.1.3. Alternative assessment tasks. To obtain the alternative assessment scores of the participants in the experimental group, the researcher designed several tasks for each language skill course. These tasks were prepared considering the objectives of each skill course, syllabus and coursebook content of each skill. In this way, the participants in the experimental group were not only engaged in alternative assessment activities but also got prepared for their school exams. In addition, the integration of skills was a major concern during the task preparation process. Specifically, reading and listening materials were incorporated into writing and speaking tasks since reading and listening are receptive and covert skills, and had to be assessed through productive tasks.

The speaking tasks were prepared in parallel with the learning objectives of the course syllabus, and the content of the coursebooks to ensure the validity of the tasks. In the School of Foreign Languages, *Prism Listening/Speaking* was used as a coursebook in

listening and speaking courses. The researcher arranged the frequency of the tasks considering the duration of the module. As mentioned previously, the integration of skills was a major concern during the preparation of the tasks. Therefore, some of the tasks were prepared based on the content of the reading coursebook *Prism Reading*. The alternative assessment methods that were used to foster the speaking skills of the participants included presentations, debates, and live or recorded role-play (video shooting) tasks. The participants were not assigned a task for some weeks since each module lasted eight weeks and the participants needed time to prepare for their traditional exams as well. In addition, the researcher did not prefer to put too much pressure on the participants by implementing a task each week. While designing each task, the researcher also prepared a guideline and a rubric for grading so that the students could be informed about the purpose of the task, what they needed to do in each task, and how their performances would be assessed.

The writing tasks were also prepared considering the objectives of the course, the syllabus and the coursebook content. In the School of Foreign Languages, *Ready to Write* is used as a coursebook in writing courses. According to the writing syllabus of Module B1, the participants first learned the basic components of a paragraph. Then, they learned how to write different paragraph types such as opinion, compare/contrast, and cause/effect paragraphs. In Module B1+, they learned the basic components of an essay. After that, they learned how to write different essay types such as opinion, compare/contrast, and cause/effect essay. In the School of Foreign Languages, writing instructors generally ask their students to write about certain topics based on their own preferences or let their students choose the topics themselves. In other words, they choose the topics randomly. On the other hand, the researcher, in this study, selected the topics of paragraphs and essays based on the content of reading and listening/speaking coursebooks. For instance, the participants read a text in the reading course, and wrote a paragraph, or an essay based on the content of the text. Similarly, they listened to a lecture in the listening/speaking course, and then wrote a paragraph, or an essay based on the lecture they listened to. In short, the researcher used the reading and listening materials to create content for the writing tasks. In this way, the participants had a chance to integrate written or audio input into writing tasks. After each task, the participants conducted self- and peer assessment using a checklist and a grading rubric. They also received written and oral feedback on their performances through one-to-one conferences with their teacher. Based on the feedback they received, the students revised their writing papers and kept all of their drafts in their paper-based portfolios. In short, the participants were engaged in integrated paragraph and essay writing activities in their writing

courses. They also prepared an online magazine in groups, which was evaluated as an end-of-term project. This online magazine consisted of certain sections such as health, education, sports etc. The students were divided into groups of four or five and they were supposed to write paragraphs for the sections of the magazine that they were assigned. Their performances in this task were also evaluated through a checklist and the writing rubric. Since the researcher prepared all the writing tasks considering the course syllabus, the participants also had a chance to prepare for the school exams. Most importantly, at the beginning of the study, the researcher informed the participants about the portfolio process, self- and peer assessment procedures, and provided a grading rubric for the participants so that they could be informed about the purpose of portfolio keeping, what they needed to do during the process, and how their performances would be assessed.

Since reading and listening are covert and receptive skills, the participants' performances in these skills were assessed through speaking and writing tasks. As mentioned in the previous sections, reading, and listening materials were incorporated into speaking and writing tasks so that the participants' reading and listening performances could be assessed in an integrated and observable way. In other words, the aim of the researcher was to adopt an integrated approach through which reading and listening content could be used as a source for productive tasks. In this way, the participants had the chance to master their four language skills (i.e. reading, listening, writing, and speaking) simultaneously.

3.3.1.4. Semi-Structured interviews. In the field of social science research, there are three main types of interviews that are frequently used by researchers. The first type is called *structured interviews*. In this type of interview, there are a set of predetermined questions, which usually require an immediate response in the form of 'yes' or 'no.' As a result, neither the interviewer nor the interviewees have much space to elaborate on the issue (Berg, 2007). Secondly, researchers employ *open-ended (unstructured) interviews*. According to Gubrium and Holstein (2002), open-ended interviews provide more flexibility and space for both interviewers and interviewees in terms of planning, designing, and conducting the interview. Therefore, the interviewer can ask various questions to extend the interview and let the interviewees elaborate on the issues (Dörnyei, 2007). The third type is called the *semi-structured interview*. This type of interview is a more flexible version of the structured interview as it provides an opportunity for the interviewer to interpret and comment on the interviewee's responses (Rubin & Rubin, 2005). While implementing semi-

structured interviews, researchers tend to use a basic checklist or specific questions (i.e., research questions) which would direct the interview into relevant areas (Berg, 2007).

For the present study, semi-structured interviews were conducted with 15 voluntary participants in the experimental group. The interview questions (See Appendix B) were revised after the pilot study, and they were finalized by the researcher and his dissertation supervisor. The final version of the interview included eight questions. The interviews were conducted face-to-face and in Turkish since the participants would express their opinions and feelings better in their mother tongue. The interviews were audio-taped for transcription purposes, and then they were transcribed by the researcher manually. After that, the transcriptions were translated into English by the researcher, and another colleague also checked the translations to increase the accuracy of the translations. Thanks to these interviews, the researcher was able to gather in-depth data regarding the views and perceptions of the participants on the implementation of alternative assessment methods. In addition, he wanted to expand on the quantitative data by comparing and relating the findings of the quantitative with qualitative data.

3.3.1.5. Learning /reflective journals. In recent years, the role of reflection has been highlighted in educational settings (Moon, 2006). It is described as a form of mental processing through which people inquire about their own beliefs, attitudes, and actions critically. According to Prinsloo et al. (2011), learning diaries have been used as a reflection tool to create awareness and foster students' critical thinking skills in classrooms. Learning diaries, also known as learning journals, can serve different purposes. First of all, students are informed about their learning processes and identify their strengths and weaknesses in certain areas. Secondly, keeping a learning diary could improve the problem-solving skills of students. They could work on their weaknesses and try to find some strategies or solutions to compensate for their weaknesses. Finally, they could assess their own performances by answering some predetermined questions in a learning diary. In this way, they could track their own progress in learning (Moon, 2004; Tsang, 2003).

Throughout the treatment process, the participants in the experimental group were supposed to keep a learning/reflective journal to evaluate the process on a weekly basis. The researcher prepared a set of guiding questions (See Appendix C) for each skill course and gathered further views from another expert, who held a PhD degree in English Language Teaching program, on the content and quality of these questions. Upon receiving some suggestions, the researcher finalized the journal questions. The participants were expected

to answer these questions at the end of each week. In this way, the researcher aimed to gather rich data as much as possible so that he could evaluate the treatment process in a more accurate way.

3.4. Data Collection Process

This study aims at finding out the effect of applying alternative assessment methods on a group of Turkish EFL students' academic achievement and investigating the attitudes of these students towards learning English before and after the implementation of alternative assessment methods in a preparatory school. At the beginning of the term, the ethics committee approval document was received on the 5th of November, 2021. In addition, the necessary permission was obtained from the administration of the School of Foreign Languages, and the consents of the participants were also taken at the beginning of the process.

3.4.1. The Pilot Study

Before the main study, a pilot study was conducted in the Fall semester of the 2021/2022 academic year. This piloting process was an essential part of the study since possible problems related to the treatment process, data collection tools and data analysis procedures could be detected by the researcher. In this way, it would be possible to make the necessary revisions and modifications accordingly. The participants of the pilot study included 67 students who would be studying in English Language Teaching (ELT) and English Language and Literature (ELL) departments in the following year. They were all placed in B1 level classrooms in the first module based on their placement exam scores. There were 33 students in the experimental group while the control group included 34 participants.

Before the treatment process, the researcher prepared a number of alternative assessment tasks for each skill. He designed these tasks based on the content of the coursebooks and learning outcomes of the syllabus of each skill so that the validity of the tasks can be ensured. To improve the quality and validity of the tasks, he also obtained the views of two experts who hold a PhD degree in the ELT program. Having finalized the tasks, the researcher selected a writing rubric (See Appendix D), an oral presentation rubric (See Appendix E), and a debate rubric (See Appendix F) to assess the performances of the participants throughout the process. The researcher made minor alterations to these rubrics as a result of the feedback he received from his dissertation supervisor. The treatment process

lasted eight weeks. The participants in the experimental group received instruction through several alternative assessment methods. They were engaged in presentations, role-plays, debates, and kept a portfolio for their writing courses. For the writing course, the participants also conducted self- and peer assessment using a checklist and the writing rubric. Finally, they received oral or written corrective feedback from the teacher on their performances through one-to-one conferences.

In short, the participants' performances during the pilot study were assessed in a formative manner by means of the selected rubrics and checklists. Therefore, they had a collection of alternative assessment scores for each skill course. To increase the validity and reliability of these scores, an external rater, who had more than fifteen-year teaching experience in the School of Foreign Languages, also graded the works of the participants. Finally, semi-structured interviews were conducted with 12 voluntary participants to reveal their views and feelings regarding the treatment process.

A few points need to be highlighted regarding the treatment process. To start with, all the alternative assessment tasks during the processes were designed according to the syllabus and textbook of each skill. In this way, the participants in the experimental group had a chance to get prepared for their school exams as well. In addition, the integration of skills was a major concern during the process. Specifically, the reading and listening materials were incorporated into writing and speaking courses since reading and listening are receptive skills and had to be assessed through productive tasks.

As for the participants in the control group, they experienced a conventional instruction, and their performances were assessed through traditional assessment methods. They used the same coursebooks and followed the same syllabi as the participants in the experimental group. However, they followed the instructions of the coursebooks, and their performances were assessed through traditional tests, which mostly included multiple-choice, true/false, and gap-filling test items. In addition, they did not keep a portfolio in their writing courses, nor did they perform self- or peer assessment during their writing lessons.

At the end of the pilot study, the researcher conducted some parametric tests to find out the effectiveness of the treatment on the participants' academic achievement as the normality assumption of the data was met. The statistical analyses revealed that the total alternative assessment scores of the participants in the experimental group did not significantly differ from their total traditional test scores. In addition, there was not a significant difference between the alternative assessment scores and traditional test scores of the experimental group participants in terms of four skills (i.e., reading, listening,

speaking, and writing). Finally, the traditional test scores of the participants in the experimental group did not significantly differ from the traditional scores of the participants in the control group. The findings of the pilot study might indicate that the treatment did not make any difference in the performances of the participants in the experimental group. However, these findings might be misleading since all of the participants in the pilot study were ELT and ELL students, so they all had a solid English background to perform well in traditional exams. Therefore, they might have attained similar traditional assessment scores. On the other hand, the interview data revealed that the participants in the experimental group mostly had a positive attitude towards the use of alternative assessment methods. They also suggested some modifications to the implementation of alternative assessment methods for the future. In general, the participants agreed that the tasks were effective and motivating for them. However, they also suggested that the number of the tasks be reduced because they felt exhausted during the process from time to time.

The piloting process provided several benefits for the researcher. First of all, the quality and appropriateness of the tasks were checked. The researcher realized that the tasks were favored by most of the participants. However, the number of the tasks assigned needed to be reconsidered by the researcher since the participants complained about their workload during the interviews. Second, the researcher was able to test the validity and reliability of the assessment tools (i.e., checklists and rubrics) and found that some minor revisions were necessary to make these tools more useful. In addition, the researcher realized that some of the interview questions needed to be modified or elaborated since they caused ambiguity for some of the participants. Most importantly, the researcher and his dissertation supervisor decided that it would be useful to inquire whether applying alternative methods would have any impact on the attitudes of the participants towards learning English. Therefore, they started searching for an attitude scale to be used in the main study.

3.4.2. The Main Study

As for the main study, the data collection instruments were applied to 75 preparatory class students whose ages ranged between 17 and 19. These were native speakers of Turkish and were studying English as a Foreign Language. As mentioned before, the convenience sampling technique was employed for the selection of the participants since the researcher was also working in the School of Foreign Languages as a language instructor. Therefore, organizing the setting, collecting the data, and managing the treatment process was convenient for the researcher. There were two groups of participants selected for the present

study. The first group was the experimental group consisting of 38 participants, while the control group had 37 students. Both experimental and control group participants were placed in B1 classes based on the placement test administered by the School of Foreign Languages at the beginning of the term. In the present study, the treatment process lasted 16 weeks in two modules as B1 and B1+. To be more precise, the process was started on the 21st of February, 2022 and was completed on the 17th of June, 2022. At the beginning of the process, an attitude scale towards learning English (See Appendix A) was administered to both experimental and control group participants to gather data about their attitudes towards learning English prior to the treatment. The consent of the participants to contribute to the study was also taken through their signatures under a statement on the same scale.

3.4.2.1. Treatment process for the experimental group. *In the first week of the module*, the participants in the experimental group were introduced to the writing rubric (See Appendix D), the oral presentation rubric (See Appendix E), and the debate rubric (See Appendix F) so that they could be informed on how their performances would be assessed throughout the term. In other words, students were aware of what criteria they would be assessed on for each skill in advance as transparent scoring criteria is an essential feature of alternative assessment applications (Brown & Hudson, 1998). Regarding the use of rubrics, Solomon (2002) pointed out that they were the tools employed to describe the performance behaviors of students based on predetermined criteria and provided evidence of achievement. The participants were also asked to get a portfolio file in which they would keep all their written works and learning journals (See Appendix E) throughout the modules. Richards and Schmidt (2010) define a portfolio as "a purposeful collection of work that provides information about someone's efforts, progress or achievement in a particular context. It is learning as well as an assessment tool" (p. 443). That is why the participants were required to possess a file to keep recordings of their performance throughout the terms.

According to the syllabus, *for the first week's reading lesson*, the students were supposed to read a text titled *Endangered Species*. Having read the text and completed the traditional exercises such as multiple-choice and gap-filling questions in the book, the participants in the experimental group were asked to prepare a PowerPoint presentation about an endangered species to be presented in the following speaking lesson. While preparing the presentations, they had to reread the text in their reading books with a critical approach this time and learned about what to include and how to integrate certain sections (i.e., causes of extinction, ways to protect endangered species etc.) and the related

vocabulary they have learned into their presentations by using their macro and micro skills in reading (Brown,2007). In other words, they were asked to move from reading to produce the language in spoken form. In the speaking lesson, the students made their presentations, and their performances were assessed based on the oral presentation rubric. During the presentations, the other students were supposed to take notes and ask some questions to the presenter afterwards. In this way, a lot of student interaction was observed during the speaking lessons. In this way, it was hoped that the participants' speaking skills, presentations skills, and communicative strategies would improve. *As for the listening lesson*, the students were supposed to listen to a debate about *Using Animals for Work*. Having done the exercises in the book, the experimental group participants were engaged in a writing activity this time. During the writing lesson, they learned how to write a well-organized paragraph that has a topic sentence, supporting sentences and a concluding sentence. As for the topic selection, they were supposed to respond to the debate on *Reasons for Using Animals for Work* in written form. To do so, they had to write a paragraph and reflect their opinions on the issue of whether animals should be used for work or not by referring to the debate they had listened. In this way, listening and writing lessons were also conducted in an integrated way. Regarding this issue, Huerta-Macias (2002) states that alternative assessment evaluates to what extent students can integrate and produce what they have been taught. Finally, the students' writing performances were assessed through the writing rubric.

In the third week of the module, students read a text titled *Traffic Congestion*. Having read the text and carried out the related tasks in the book, the participants in the experimental group were asked to prepare a PowerPoint presentation on the causes or effects of traffic congestion in their hometowns and come up with some solutions to reduce this problem. They seemed motivated since the task was related to their real-life situations (Huerta-Macias, 2002). In order to do so, they were also asked to refer to the text they had read and see if there were any similar or different points between their hometowns and the places mentioned in the article in terms of traffic congestion issues. They also had a chance to use the new vocabulary they learned in the text actively. In this way, they were able to gain a deeper understanding of the text. *Starting from the third week*, during the writing lessons, the students were also trained on how to assess a paragraph using the checklist and writing rubric. It was obvious that they had never heard of what a rubric was and how it was used. Therefore, the students were asked to mark several sample papers with their peers and by themselves using the writing rubric until they had an adequate understanding of how to use

it. At first, the students were confused and struggled a lot. Naturally, there were huge score differences among students' grading. However, they got better in time as they were provided feedback on how to use the rubric properly throughout their writing lessons. The students were also informed that they would perform self-assessment and peer assessment using the checklist in their writing coursebooks and the writing rubric in their writing classes as well as receiving corrective feedback from their teachers. Moqbel (2015) defines self-assessment as a process in which students are expected to assess their performance against certain standards, and it also refers to procedures through which students themselves evaluate their language skills and knowledge. On the other hand, peer-assessment enables students to develop their collaborative skills as well as getting more opportunities to learn from each other. In this way, they could be more aware of the objectives of the course as well (Moqbel, 2015). Accordingly, the participants in the study were able to detect their strengths and weaknesses in certain areas (i.e., content, organization, spelling etc.) and tracked their own progress through self-assessment and peer-assessment. *As for the listening lesson*, the students were supposed to listen to a radio program about *Phobias: Fear of Flying*. Having done the exercises in the book, the students were asked to choose a type of phobia and write a cause paragraph on it. While writing the paragraph, they had to refer to the radio program that they had listened to before and learned what sections to include (i.e., definition of the phobia, what could cause it, and how to overcome it) to include into their writings by dwelling on their micro and macro skills in listening (Brown, 2007). The students performed self- and peer assessment using the checklist on their coursebook and the writing rubric provided by the teacher. In this way, they were able to reflect on their writing and discuss their strengths and weaknesses with their peers. Finally, their performances were assessed through the writing rubric by the teacher. Starting from the third week, the students were asked to reflect on the tasks and practices that they had been engaged in each week by filling in their learning journals (See Appendix C). Student journals are also regarded as an alternative assessment technique. However, these are less structured than portfolios and the main aim is to maintain a steady communication and build a strong rapport between the student and the teacher (Sheppard & Stoller, 1995; Ur, 1996). In the present study, the participants' views and feelings regarding the treatment process were gathered through their journals so that necessary modifications could be made on the treatment procedure if necessary. In addition, thanks to these journals, individual needs of the students were detected, and the researcher dealt with these needs through one-to-one conferences with the students throughout the process.

In the fifth week of the module, the students read a text titled *Customs around the world*. After they had done the exercises in the book, the students were asked to shoot a video in which they were supposed to include different customs and traditions around the world. Before shooting their videos, they were advised to read the text again and write a short script about how different customs could cause problems or lead to misunderstandings or frustrations in people's lives. In their videos, they acted out by focusing on these issues. The students worked in groups and prepared their videos by the deadline set by the teacher. In this way, they had a chance to improve their higher-order skills such as collaborating and thinking critically since these skills are highly important in today's educational settings. The whole class watched some of the videos during the speaking lesson and provided feedback for their classmates as well. Their performances were assessed through a checklist that was prepared by the teacher and introduced to them beforehand. As for the writing lesson, they were asked to write a paragraph comparing two different countries in terms of their customs and traditions. Since they had already read about how countries differ from one another in certain areas (i.e., dress code, table manners, punctuality, etc.), and practiced some vocabulary items on the issue, they were able to come up with a lot of ideas for their paragraphs. Having written their paragraphs, the students assessed their own papers using the checklist and writing rubric. Then they assessed their peers' work and discussed their grading criteria with each other referring to the checklist on their coursebook and the given writing rubric. Thanks to the rubric, they were able to detect their strengths and weaknesses in certain areas with their peers. As the last step, the students had feedback sessions with their teacher during one-to-one conferences. As time passed, it was observed that students' self-assessment and peer-assessment scores were similar to the scores given by their teacher.

In the seventh week of the module, a debate was organized, and students were put into groups by the teacher so that a balance in terms of proficiency level was maintained among the groups. Each group was assigned a topic that had been covered in the reading or listening/speaking lessons previously. The groups had enough time to revise the topics in their reading or listening/speaking books and get prepared for the debate. They were also reminded that their debate performances would be assessed through the debate rubric (See Appendix F). During the debate, each group was supposed to listen to the other group and take notes so that they could refute their opponents' ideas later. At the end of the debate, there was also a question/answer session where the audience asked questions to each group related to the topic they had discussed. In this way, the whole class participated in the debate activity and a larger amount of student interaction was observed throughout the lesson. In

short, the participants were engaged in several integrated tasks during Module B1, and their performances were assessed steadily throughout the process based on certain rubrics and checklists. In addition, their views and feelings regarding the process were gathered through their learning journals. The tasks assigned to the participants during Module B1 are shown in Table 3.2.

Table 3.2. *Schedule of the Assigned Tasks for the Experimental Group (Module B1)*

<i>Week</i>	<i>Date</i>	<i>Topic</i>	<i>Tasks Assigned</i>
Week 1	21-25 February 2022	Endangered Species Using animals for work	Oral presentation Integrated writing task
Week 3	07-11 March 2022	Traffic congestion Phobias	Oral Presentation Integrated writing task
Week 5	21-25 March 2022	Customs around the World Comparing different customs	Shooting a Video Integrated writing task
Week 7	4-8 April 2022	Diverse topics from coursebooks	Debate

At the end of module B1, both experimental and control group participants took the final exam. The students whose general points of average were 70 and above were considered successful and had the right to start B1+ module. These students were placed in experimental and control groups with their previous classmates and teachers. In other words, the treatment process continued under the same circumstances during B1+ module. As the experimental group participants had already got familiar with alternative assessment techniques and tools during B1 module, they did not have any difficulty in adapting to the process in B1+ module. The tasks assigned to the participants during Module B1+ are presented in Table 3.3.

Table 3.3. *Schedule of the Assigned Tasks for the Experimental Group (Module B1+)*

<i>Week</i>	<i>Date</i>	<i>Topic</i>	<i>Tasks Assigned</i>
Week 1	25-29 April 2022	Globalization Effects of globalization	Oral presentation Integrated writing task
Week 3	9-13 May 2022	Distance vs. face-to-face Learning Pandemics	Oral Presentation Integrated writing task
Week 5	23-27 May 2022	Diverse topics from coursebooks	Debate
Week 7	6-10 June 2022	Topics assigned by the teacher	Online Magazine

To sum up, the experimental group participants were engaged in alternative assessment techniques such as presentations, debates, and video shooting assignments, and kept a portfolio for their writing courses. In addition, their performances were also assessed by means of certain rubrics and checklists during self and peer assessment as well as one-to-one conference sessions with the teacher in a formative manner. Therefore, they were able to receive continuous constructive feedback and support from their peers as well as the teacher and had a collection of alternative assessment scores for each skill. In addition, their views and feelings regarding the learning process were gathered through their learning journals so that their individual needs could be handled by the teacher.

A few points need to be highlighted regarding the treatment process. To start with, all the alternative assessment tasks during the processes were designed according to the learning objectives of the syllabus and textbook of each skill. In this way, the experimental group participants were not only engaged in alternative assessment activities but also got prepared for their school exams. In addition, the integration of skills was a major concern during the process. Specifically, the reading and listening lessons were integrated into writing and speaking lessons through the pre-designed tasks since reading and listening are covert and receptive skills and had to be assessed through productive tasks. In other words, reading and listening materials provided the necessary input or content for the speaking and writing tasks. Finally, the students were actively engaged in the treatment and assessment processes through self-assessment, peer assessment, and learning journals.

The main purpose of implementing alternative assessment methods was to encourage the participants to use the language communicatively both inside and outside the classroom. In addition, the researcher aimed to provide an opportunity for the participants to integrate four language skills through meaningful tasks. In other words, the participants had

a chance to transform the receptive input into creative meaningful output through productive tasks. Most importantly, the researcher attempted to improve problem-solving skills of the participants by assigning them both individual and group tasks. In this way, more interaction and collaboration were observed in the classroom. Since the tasks were designed based on the coursebook and syllabus of each skill course, the participants were able to get prepared for their school exams as well. Finally, based on the general observations, adopting a learner-centered approach was effective in increasing the motivation level of the participants as well as promoting learner autonomy.

3.4.2.2. The process for the control group. As for the control group participants, they experienced a conventional teaching and assessment process. They used the same textbooks and followed the same syllabi with the experimental group participants. *In the reading lessons*, they read the texts in their textbooks and did the follow-up exercises, which included multiple-choice, true-false and gap-filling questions. *In the listening lessons*, they listened to the tracks and completed the exercises in their textbooks. These exercises included multiple choice questions and note-taking parts. *In the speaking lesson*, they had discussions on the topics of the units as a whole class and sometimes in groups. They also made presentations from time to time. In short, they were also engaged in communicative activities like the participants in the experimental group. However, their performances were not assessed through any type of rubrics determined beforehand. Therefore, the students did not know on what criteria they were being assessed. In other words, they received scores through a holistic assessment by their teachers. *In the writing lessons*, they wrote about diverse topics according to their interests or their teachers' preferences. The students received continuous written or oral feedback from their teachers throughout the writing lessons so that they could determine their strengths and weaknesses in certain sections in writing. However, they did not perform any type of self-assessment or peer assessment throughout the process. In addition, they did not keep their writing activities in a portfolio throughout the process. Finally, the participants in the control group were assigned performance worksheets and took a number of quizzes for each skill by their teachers. However, these included the same traditional types of questions. In short, the participants in the control group also had a busy schedule and completed several tasks for each skill during the process. They were also assessed several times through traditional tests during the process.

All in all, it could be stated that experimental group participants were engaged in alternative assessment techniques and tasks where four skills (i.e., reading, listening, speaking, and writing) were integrated, and their performances were assessed in a formative manner through predetermined rubrics and checklists throughout the process. Since they performed self and peer assessments and had one-to-one conferences with their teacher, they had the opportunity to assess their own performances and receive continuous feedback both from their teachers and peers. Furthermore, their views and feelings regarding the process were gathered through learning journals so that the teaching and assessing procedures could be modified if necessary. On the other hand, the participants in the control group experienced conventional instruction during the process. For each skill lesson, they completed the existing exercises in their textbooks and did the worksheets assigned as homework. In the speaking lesson, they completed some communicative tasks, but unlike the participants in the experimental group, their performances were not assessed through any predetermined rubrics. In the writing lessons, they wrote paragraphs and essays like the participants in the experimental group. However, they did not keep a portfolio or conduct any self- and peer-assessment in a systematic way in their writing lessons. Finally, as mentioned above, the participants in the experimental group were involved in predesigned integrated tasks in which they were supposed to incorporate the reading and listening input into productive tasks either in written or spoken form. On the other hand, the integration of skills was not a concern for the tasks assigned to the participants in the control group.

At the end of the treatment process, the same attitude scale towards learning English was administered to both experimental and control group participants to see whether there was a statistically significant difference between their pretest and post test scores based on the scale data. In addition, alternative assessment scores of the experimental group participants were gathered for each skill. To ensure the inter-rater reliability of these scores, an external rater, who held a PhD degree in the ELT field and worked in the School of Foreign Languages, also assessed the data throughout the process. In addition to the alternative assessment scores, traditional exam scores of both experimental and control group participants were obtained from the testing office of the School of Foreign Languages. This was necessary to compare the alternative and traditional assessment scores of these participants in the experimental group. In addition, as the research questions of the study indicate, the traditional assessment scores of both experimental and control group participants would be compared to see if there was a significant difference between the academic achievement of both groups. Finally, semi-structured interviews were conducted

with 15 voluntary experimental group participants to gather their views and feelings on the implementation and effectiveness of alternative assessment applications.

3.5. Data Analysis

The present study employed a mixed-method sequential explanatory research design. The quantitative data included scale data as well as traditional and alternative assessment scores of the participants. On the other hand, the qualitative data were collected through semi-structured interviews and reflective learning diaries (journals). To analyze the quantitative data, the researcher employed the Statistical Package for Social Sciences (SPSS 20), and a number of statistical analyses were performed for the corresponding research questions. The qualitative data were subject to content analysis and the emerging themes were identified with the help of another external rater, who held a PhD degree in the ELT field.

First, the normality of the data was checked so that the appropriate test could be selected for data analysis. The researcher applied Kolmogorov-Smirnov test to check the normality assumption since the number of participants is above 50 (Büyüköztürk, 2013). The analyses of the Kolmogorov-Smirnov test indicated that the data followed a normal distribution. The analysis of Skewness and Kurtosis values and histograms also confirmed this finding. Therefore, the researcher conducted parametric tests to analyze the rest of the data. To analyze the scale data, paired-samples t-test was conducted to see whether there was a significant difference between the pretest and posttest scores of experimental and control group participants within each group. Another paired-samples t-test was conducted to compare the overall alternative assessment scores and traditional assessment scores of the participants within the experimental group. To compare the alternative and traditional assessment scores of the participants in terms of four skills within the experimental group, the researcher again conducted a number of paired-samples t-tests.

To see whether there was a significant difference between the overall traditional assessment scores of the participants in the experimental group and the control group, an independent-samples t-test was conducted. Finally, the researcher conducted a number of independent-samples t-test to compare the traditional assessment scores of the experimental and control group participants in terms of four skills. On the other hand, the qualitative data collected through semi-structured interviews and learning diaries (journals) were analyzed through content analysis. During this content analysis, the researcher followed a zigzag

pattern, moved back and forth by constantly comparing the flow of the coming data until certain codes and themes emerged (Cresswell, 2007). To ensure the reliability of the content analysis, the researcher received support from another expert who held a PhD degree in the ELT field. A summary of the research design, participants and procedures is presented in Table 3.4.

Table 3.4. Summary of the Research Design, Participants and Procedures

Research Design	Mixed-Methods, both quantitative and qualitative
Sample Selection	Convenience Sampling, Purposeful Sampling (For the Interview)
Participants	Experimental group consisted of 38 students while control group included 37 students
Data Collection Tools	Attitude Scale towards Learning English, Alternative Assessment Tasks and Traditional Exams, Reflective Journals, Face-to-face Semi-structured Interviews
Data Analysis Tools	SPSS 20 (for the analysis of scale data and assessment scores), content analysis (for the interviews and reflective journals)
Treatment Practice	Application of certain alternative assessment techniques and tasks through the integration of four language skills (i.e., reading, listening, speaking, and writing)
Time and Duration	From 21 st February 2022 to 17 th June 2022 (i.e., B1 and B1+ Modules)

CHAPTER 4: FINDINGS

In this section, the findings of the quantitative and qualitative analyses are presented with respect to the research questions. The quantitative analyses were conducted by means of SPSS 20 statistical software, and the differences were accepted as statistically significant when p value is ≤ 0.05 . On the other hand, the qualitative data collected through interviews were subjected to content analysis so that the emerging themes could be identified.

Before answering the research questions, the normality assumption of the data was checked through the Kolmogorov-Smirnov test. According to Büyüköztürk (2013), if the number of participants is above 50, the Kolmogorov-Smirnov test is applied to check the normality assumption. In addition, Skewness and Kurtosis scores were checked to determine whether the data had a normal distribution. Since Skewness and Kurtosis values were between -2 and +2, the scale data followed a normal distribution (Tabachnick & Fidell, 2013). As Table 4.1. shows, the data followed a normal distribution. Therefore, a number of parametric tests were applied.

Table 4.1. *Normality Test Results of the Participants*

	<i>N</i>	<i>M</i>	<i>sd</i>	Skewness	Kurtosis	<i>Sig.</i>
Control	40	79.42	7.33	.159	-.452	.82
Exper.	39	72.87	8.27	-.483	.456	.20

Table 4.2. *Comparison of the Participants' Placement Test Scores*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Control	40	79.42	7.33	-.25	77	.80
	Exper.	39	72.87	8.27			

An independent samples t-test was applied to compare the placement test scores of the participants so that their existing English proficiency could be determined. Table 4.2. shows that there is not a statistically significant difference between the experimental and control group participants in terms of their English proficiency level prior to the treatment ($p = .80$), which indicates that the participants in both groups possessed similar linguistic proficiency prior to the intervention.

4.1. RQ1. Is There a Statistically Significant Difference between the Overall Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group?

To answer this research question, a paired samples t-test was conducted. The purpose of the test was to compare the overall alternative and traditional assessment scores of the participants in the experimental group.

Table 4.3. *Comparison of the Overall Alternative and Traditional Assessment Scores of the Participants in the Experimental Group*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Altern.	32	83.46	10.19	5.33	31	.00
	Tradit.	32	72.34	6.06			

As Table 4.3. indicates, there is a statistically significant difference between the overall alternative assessment scores and the traditional assessment scores of the participants in the experimental group ($p = .00$). In other words, the participants in the experimental group performed much better during alternative assessment applications and obtained significantly higher scores as a result of their performances. The reason for this difference could stem from the fact that the reading and listening scores of the participants were obtained through integrated tasks during alternative assessment applications in a formative manner. On the other hand, their traditional reading and listening scores were gathered in a summative manner from tests that included multiple-choice and gap-filling test items. Therefore, their total alternative and traditional assessment scores could have been affected by this situation.

4.1.1. RQ1a. Is There a Statistically Significant Difference between the Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Writing Skills?

Another paired-samples t-test was conducted to compare the alternative and traditional scores of the participants in the experimental group in terms of writing skills.

Table 4.4. *Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Writing Skills*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Altern.	32	88.46	8.81	1.50	31	.14
	Tradit.	32	85.46	9.78			

As indicated in Table 4.4. , it is clear that there is not a statistically significant difference between the alternative and traditional assessment scores of the participants in the experimental group in terms of writing skills ($p= .14$) This result suggests that the participants attained similar scores in their traditional tests and alternative assessment tasks in terms of writing skill. The reason for this similarity could stem from the fact that the participants were engaged in so many writing tasks during their alternative assessment applications that they performed well in their traditional writing exam as well. In addition, the items in both examinations were similar to each other because they wrote essays on similar topics.

4.1.2. RQ1b. Is There a Statistically Significant Difference between the Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Reading Skills?

Another paired samples t-test was applied to compare the alternative and traditional scores of the participants in the experimental group in terms of reading skills.

Table 4.5. *Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Reading Skills*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Altern.	32	80.93	18.02	2.20	31	.03
	Tradit.	32	72.50	13.50			

According to Table 4.5, it is clear that there is a statistically significant difference between the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of reading skills ($p= .03$). The reason for this difference could stem from the fact that the participants reading scores were attained through integrated tasks during alternative assessment applications in a formative manner. On the other hand, their traditional scores were obtained from their standardized multiple-choice reading tests which were applied in a summative manner.

4.1.3. RQ1c. Is There a Statistically Significant Difference between the Alternative Assessment scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Listening Skills?

To answer this research question, another paired samples t-test was conducted to compare the alternative and traditional assessment scores of the participants in the experimental group in terms of listening skills.

Table 4.6. *Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Listening Skills*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Altern.	32	80.37	15.48	5.26	31	.00
	Tradit.	32	59.56	15.16			

As seen in Table 4.6., there is a statistically significant difference between the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of listening skills ($p = .00$). Like reading skill, listening is also a covert and receptive skill. Therefore, the participants' alternative assessment scores were attained through integrated tasks in a formative manner. On the other hand, the participants' traditional scores were gathered from timed, multiple-choice listening tests that were applied in a summative manner. As a result, the participants could have obtained different scores from these two different assessment methods.

4.1.4. RQ1d. Is There a Statistically Significant Difference between the Alternative Assessment Scores and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Speaking Skills?

To compare the alternative and traditional assessment scores of the experimental group participants in terms of speaking skills, another paired samples t-test was conducted.

Table 4.7. *Comparison of the Alternative and Traditional Assessment Scores of the Participants in the Experimental Group in terms of Speaking Skills*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Altern.	32	87.40	12.32	.45	31	.65
	Tradit.	32	86.06	11.31			

As indicated in Table 4.7., there is not statistically a significant difference between the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of speaking skills ($p = .65$). In other words, the participants attained very similar scores from their alternative assessment speaking tasks and traditional speaking exams. The reason for this similarity could stem from the fact that both alternative assessment and traditional assessment procedures included tasks that made the participants use their communicative skills in similar ways. In addition, the participants were engaged in

so many communicative tasks throughout the term that they were able to perform well in their traditional speaking exams as well.

4.2. RQ2. Is There a Statistically Significant Difference between the Overall Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group?

To answer this research question, an independent samples t-test was conducted. The purpose of this test was to see if there was a statistically significant difference between the overall traditional scores of the participants in the experimental and control groups.

Table 4.8. *Comparison of the Overall Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Control	34	65.85	7.93	-3.71	64	.00
	Exper.	32	72.34	6.06			

As seen in Table 4.8., there is a statistically significant difference between the overall traditional assessment scores of the participants in the experimental group and control group ($p = .00$). This finding could indicate that alternative assessment applications had a positive effect on the academic achievement of the participants in the experimental group since they attained significantly higher scores in their traditional exams compared to the participants in the control group.

4.2.1. RQ2a. Is There a Statistically Significant Difference between the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Writing Skills?

Table 4.9. *Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Writing Skills*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Control	34	83.17	10.61	-.91	64	.36
	Exper.	32	85.46	9.78			

As shown in Table 4.9., there is not a statistically significant difference between the traditional assessment scores of the participants in the experimental group and control group

in terms of writing skills ($p= .36$). The reason for this similarity could stem from the fact that the participants in both experimental and control groups were engaged in so many writing tasks during the writing classes that they attained more or less the same scores in their traditional writing exam as well.

4.2.2. RQ2b. Is There a Significant Difference between the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Reading Skills?

Table 4.10. *Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Reading Skills*

	Group	N	Mean	sd	t	df	p
Pair	Control	34	64.82	14.63	-2.21	64	.03
	Exper.	32	72.50	13.50			

Table 4.10 indicates that there is a significant difference between the traditional assessment scores of the participants in the experimental group and control group in terms of reading skills ($p= .03$). The participants in the experimental group were engaged in integrated tasks for which they had to read several articles in their coursebooks. During these readings, they performed several micro and macro skills such as deducing the meaning and inferencing in reading as well. Then they were supposed to complete some productive tasks such as writing an essay, preparing a presentation, or shooting a video based on the articles they had read. Therefore, they both improved their productive and receptive skills during alternative assessment applications.

4.2.3. RQ2c. Is There a Statistically Significant Difference between the Traditional Assessment scores of the Participants in the Experimental Group and the Control Group in terms of Listening Skills?

Table 4.11. *Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Listening Skills*

	Group	N	Mean	sd	t	df	p
Pair	Control	34	45.52	16.36	-3.60	64	.00
	Exper.	32	59.56	15.16			

According to Table 4.11, it could be stated that there is a statistically significant difference between the traditional assessment scores of the participants in the experimental group and control group in terms of listening skills ($p = .00$). Despite this difference, the participants in both groups could not attain very high scores in their listening exams since it appears to be a really challenging skill to master. Still, as they were exposed to a considerable amount of target language during the alternative assessment tasks, the participants in the experimental group improved their listening skills to some extent and performed better in the traditional listening exam compared to their counterparts in the control group.

4.2.4. RQ2d. Is There a Statistically Significant Difference Between the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in Terms of Speaking Skills?

Table 4.12. *Comparison of the Traditional Assessment Scores of the Participants in the Experimental Group and the Control Group in terms of Speaking Skills*

	Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair	Control	34	78.82	12.49	-2.46	64	.01
	Exper.	32	86.06	11.31			

Table 4.12. clearly shows that there is a statistically significant difference between the traditional assessment scores of the participants in the experimental group and control group in terms of speaking skills ($p = .01$). As mentioned before, the participants in the experimental group were involved in several speaking tasks during which they had ample opportunities to use the target language. Therefore, they improved their speaking skill dramatically and thus performed better in their traditional speaking exam as well.

4.3. RQ3. What are the Participants' Attitudes towards Learning English Before and After the Implementation of Alternative Assessment Applications?

To answer RQ3, an attitude scale towards learning English (See Appendix A) was applied to both experimental and control groups in Turkish before and after the treatment process. The attitude scale was developed by Tulgar (2018). It is a 5-point Likert Scale consisting of 26 items with options ranging from "Completely Agree (5)" to "Completely Disagree (1)". The scale was originally applied to 370 participants, and the analyses of internal consistency, half-split and test-retest reliability were conducted to check

the reliability of the scale. The factor structure of the scale was analyzed through Exploratory Factor Analysis and four dimensions explaining the 64.55% of the total variance were obtained and the model fit indices were at a good level ($\chi^2/SD=2,93$, $RMSEA=.069$, $SRMR=.055$, $RMR=.046$, $NFI=.97$, $NNFI=.98$, $CFI=.98$, $IFI=.98$, $RFI=.97$, $AGFI=.86$, $GFI=.86$). The reliability coefficient of the whole scale and the sub-dimensions were found to be at a good level and Cronbach's coefficient in general was found as .87. As a result of the calculations, it was determined that a score between 26-60 refers to a *low attitude*, 61-95 *medium attitude*, and 96-130 *high attitude*. Finally, in the present study, Cronbach's coefficient in general was found as .85.

Before analyzing the scale data, the normality assumption of the data was checked through the Kolmogrow-Smirnow test. According to Büyüköztürk (2013), if the number of participants is above 50, the Kolmogorov-Smirnov test is applied to check the normality assumption. In addition, Skewness and Kurtosis scores and Histograms were checked to determine whether the data had a normal distribution. Since Skewness and Kurtosis values were between -2 and +2, the scale data followed a normal distribution (Tabachnick & Fidell, 2013). As indicated in Table 4.13., the data followed a normal distribution.

Table 4.13. *Normality Test Results of the Attitude Scale*

	<i>N</i>	<i>M</i>	<i>sd</i>	Skewness	Kurtosis	<i>Sig.</i>
Pretotal	75	106.64	9.08	-.06	.21	.94

The Histogram also indicates a normal distribution of the data. Therefore, a number of parametric tests were applied.

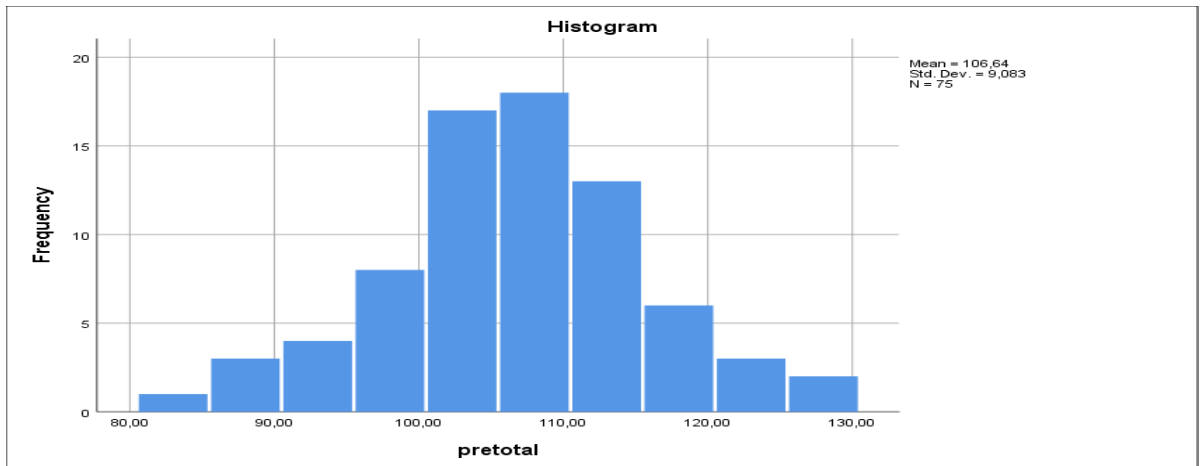


Figure.4.1 Histogram for the normality of the attitude scale

4.3.1. RQ3a. Is There a Statistically Significant Difference between the Pretest and Posttest Scores of the Participants in the Control Group in terms of Their Attitudes towards Learning English?

In order to answer the research questions, a paired samples t-test was conducted on the participants' responses in the control group to see whether there was a statistically significant difference between their pre and posttest scores in terms of their attitudes towards learning English. The descriptive statistics of the paired samples t-test for the control group participants are presented in Table 4.14.

Table 4.14. Comparison of the Pretest and Posttest Scores of the Participants in the Control Group on the Attitude Scale

	Group	N	Mean	sd	t	df	p
Pair	Pretest	38	107.39	9.07	4.86	37	.00
	Posttest		97.00	11.51			

Table 4.14 indicates that there is a statistically significant difference between the pretest and posttest scores of the participants in the control group based on the scale data ($p = .00$). This suggests that although they had a relatively high level of positive attitude towards learning English at the beginning of the term, the participants in the control group demonstrated a lower level of attitude by the end of the term. The following table presents the descriptive statistics of the control group participants for each item.

Table 4.15. *Descriptive Statistics of the Participants in the Control Group on the Attitude Scale (Item-based)*

	Items		<i>M</i>	<i>Sd.</i>	<i>t</i>	<i>p</i>
1	Learning English is a pleasure for me.	Pre	4.39	.595	1.614	.115
		Post	4.13	.844		
2	Learning English is important for me.	Pre	4.84	.370	1.152	.257
		Post	4.71	.515		
3	I think that learning English increases my general culture.	Pre	4.68	.471	2.438	.020*
		Post	4.39	.755		
4	Learning English is fun for me.	Pre	4.08	.632	2.313	.026*
		Post	3.76	.852		
5	I think that every student should learn English.	Pre	4.05	1.064	-.580	.565
		Post	4.18	.896		
6	I look forward to English lessons.	Pre	3.55	.645	2.499	.017*
		Post	3.16	.855		
7	I think that learning English helps me keep up with the globalizing world.	Pre	4.68	.574	2.321	.026*
		Post	4.34	.745		
8	I think that learning English contributes to my awareness of the similarities/differences between languages.	Pre	4.18	.982	.393	.697
		Post	4.11	.894		
9	I don't have difficulty in learning English.	Pre	3.03	.944	.669	.507
		Post	2.89	.863		
10	I think the "Writing Skills Course" improves my writing skills.	Pre	4.08	.818	-.758	.453
		Post	4.21	.777		
11	I think the "Speaking Skills Course" improves my speaking skills.	Pre	3.68	.962	1.954	.058
		Post	3.26	1.107		
12	I would take English preparatory education even if it is optional.	Pre	4.76	.490	1.389	.173
		Post	4.55	.795		
13	I think that learning English improves my perspective towards the mother tongue.	Pre	4.16	.916	1.303	.201
		Post	3.89	1.060		
14	I think that learning English is an opportunity to go abroad.	Pre	4.82	.457	2.401	.022*
		Post	4.47	.862		
15	I think that "Core Language Course" has increased my knowledge of grammar.	Pre	4.42	.722	.329	.744
		Post	4.37	.675		
16	I enjoy reviewing topics while learning English.	Pre	3.79	.843	2.343	.025*
		Post	3.32	1.042		
17	I think that learning English increases my self-confidence.	Pre	4.42	.919	1.705	.097
		Post	4.11	.798		
18	I think that learning English increases my knowledge of the features of language use.	Pre	4.34	.669	1.348	.186
		Post	4.13	.777		
19	I try to watch foreign TV series/movies in English to improve my English proficiency.	Pre	4.61	.679	1.549	.130
		Post	4.37	.819		

(see next page)

Table 4.15. Continued. *Descriptive Statistics of the Participants in the Control Group on the Attitude Scale (Item-based)*

	Items		<i>M</i>	<i>Sd.</i>	<i>t</i>	<i>p</i>
20	I think the “Reading Course” improves my reading skills.	Pre	4.18	.801	1.907	.064
		Post	3.74	1.107		
21	English lesson hours at the preparatory school should be increased.	Pre	2.11	.894	-.347	.731
		Post	2.18	1.087		
22	I think that learning English improves my ability to express myself.	Pre	4.29	.732	2.018	.051
		Post	3.92	.912		
23	I try to think in English to improve my English proficiency.	Pre	3.92	.850	1.022	.314
		Post	3.74	.795		
24	I think the “Listening Skills Course” improves my listening skills.	Pre	4.00	.771	2.939	.006*
		Post	3.32	1.210		
25	To improve my English proficiency, I try to write down the expressions I think of in English.	Pre	3.71	.956	-.839	.407
		Post	3.89	.894		
26	I think that learning English improves my perspective towards foreign languages.	Pre	4.61	.495	2.772	.009*
		Post	4.24	.714		

*The p values with an asterisk indicate a significant difference

Table 4.15 indicates that a significant difference was observed between the pretest and posttest scores of eight items. Items 3, 4, 6, 7, 14, 16, 24, and 26 were found to have significant differences between their pretest and posttest scores, with pretest scores being higher.

4.3.2. RQ3b. Is There a Statistically Significant Difference between the Pretest and Posttest Scores of the Participants in the Experimental Group in terms of Their Attitudes towards Learning English ?

To answer the research questions, another paired samples t-test was conducted on the participants’ responses in the experimental group to see whether there was a statistically significant difference between their pre and posttest scores regarding their attitudes towards learning English. The descriptive statistics of the paired samples t-test for the experimental group participants are presented in Table 4.16.

Table 4.16. *Comparison of the Pretest and Posttest Scores of the Participants in the Experimental Group on the Attitude Scale*

Group	<i>N</i>	Mean	<i>sd</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair Pretest	37	105.86	9.15	1.82	36	.07
Posttest		102.37	7.71			

As indicated in Table 4.16., there is not a statistically significant difference between the pretest and posttest scores of the participants in the experimental group ($p = .07$) in terms of their attitudes towards learning English. This finding could indicate that the participants in the experimental group had a relatively high level of positive attitude towards learning English at the beginning of the term and they maintained their high level of positive attitudes at the end of the term as well.

Table 4.17. *Descriptive Statistics of the Participants in the Experimental Group on the Attitude Scale (Item-based)*

Items		<i>M</i>	<i>Sd.</i>	<i>t</i>	<i>p</i>
1 Learning English is a pleasure for me.	Pre	4.32	.626	.190	.850
	Post	4.30	.618		
2 Learning English is important for me.	Pre	4.89	.315	.000	1.000
	Post	4.89	.315		
3 I think that learning English increases my general culture.	Pre	4.59	.551	.388	.701
	Post	4.54	.605		
4 Learning English is fun for me.	Pre	4.11	.809	-.442	.661
	Post	4.19	.660		
5 I think that every student should learn English.	Pre	4.00	.972	-.852	.400
	Post	4.19	.845		
6 I look forward to English lessons.	Pre	3.73	.932	-1.743	.090
	Post	4.05	.621		
7 I think that learning English helps me keep up with the globalizing world.	Pre	4.73	.560	2.252	.031*
	Post	4.38	.681		
8 I think that learning English contributes to my awareness of the similarities/differences between languages.	Pre	4.19	.701	.702	.487
	Post	4.08	.759		
9 I don't have difficulty in learning English.	Pre	2.97	1.013	-.865	.393
	Post	3.16	.986		
10 I think the "Writing Skills Course" improves my writing skills.	Pre	4.16	.688	-4.789	.000*
	Post	4.81	.397		
11 I think the "Speaking Skills Course" improves my speaking skills.	Pre	3.54	.1145	-2.317	.026*
	Post	4.19	.908		
12 I would take English preparatory education even if it is optional.	Pre	4.78	.479	1.550	.130
	Post	4.54	.931		

(see next page)

Table 4.17. Continued. *Descriptive Statistics of the Participants in the Experimental Group on the Attitude Scale (Item-based)*

	Items		<i>M</i>	<i>Sd.</i>	<i>t</i>	<i>p</i>
13	I think that learning English improves my perspective towards the mother tongue.	Pre	3.95	1.053	-.544	.590
		Post	4.08	.829		
14	I think that learning English is an opportunity to go abroad.	Pre	4.73	.508	.215	.831
		Post	4.70	.463		
15	I think that “Core Language Course” has increased my knowledge of grammar.	Pre	4.38	.639	.1641	.110
		Post	4.08	1.038		
16	I enjoy reviewing topics while learning English.	Pre	3.41	.927	-.393	.697
		Post	3.49	.804		
17	I think that learning English increases my self-confidence.	Pre	4.41	.686	.433	.668
		Post	4.32	.818		
18	I think that learning English increases my knowledge of the features of language use.	Pre	4.16	.727	-.329	.744
		Post	4.22	.712		
19	I try to watch foreign TV series/movies in English to improve my English proficiency.	Pre	4.30	.996	1.707	.097
		Post	3.97	1.040		
20	I think the “Reading Course” improves my reading skills.	Pre	4.27	.693	1.071	.291
		Post	4.05	.848		
21	English lesson hours at the preparatory school should be increased.	Pre	2.24	.925	1.222	.230
		Post	2.00	.943		
22	I think that learning English improves my ability to express myself.	Pre	4.22	.854	-.433	.668
		Post	4.30	.812		
23	I try to think in English to improve my English proficiency.	Pre	3.89	.906	-2.351	.024*
		Post	4.32	.709		
24	I think the “Listening Skills Course” improves my listening skills.	Pre	3.97	.928	1.502	.142
		Post	3.65	.949		
25	To improve my English proficiency, I try to write down the expressions I think of in English.	Pre	3.59	1.092	-1.909	.064
		Post	4.05	.880		
26	I think that learning English improves my perspective towards foreign languages.	Pre	4.32	.884	-.141	.889
		Post	4.35	.716		

*The p values with an asterisk indicate a significant difference

As the descriptive statistics on Table 4.17 indicates, there is a statistically significant difference in the pretest and posttest scores of four items. Items 7, 10, 11 and 23 were found to have significant differences in their pretest and posttest scores, with posttest scores being higher except for item 7.

4.4. Findings from the Interviews

To elaborate on the findings of the research questions and reveal the perceptions of the participants regarding alternative assessment applications, 15 students were interviewed

at the end of the treatment process. These participants were asked eight questions to find out their opinions and feelings about alternative assessment applications and how these applications affected their language learning process. The interviews were conducted in the participants' mother tongue, namely Turkish, so that they could express their opinions and feelings more freely. The interviews were then transcribed into English, and another expert with a PhD degree in the ELT field also checked the translation of the data to reach an agreement on the accuracy of the translation.

After the translation phase was finalized, the researcher started the analysis of the qualitative data following the guidelines suggested by Cresswell (2009). As the initial step, he read the data repeatedly to obtain a comprehensive understanding of the participants' responses. Then he identified certain codes based on the commonalities of the responses using a coding scheme. These codes were then reviewed several times to see whether any relationships existed among them. As the final step, the researcher identified several themes based on these relationships. To ensure the reliability of the analysis, another expert was asked to analyse the whole data following the same steps. The researcher informed this expert about the scope and aim of the study and provided detailed information on the steps of qualitative analysis (Cresswell, 2009). After the expert finished her analysis, the researcher had a meeting with the expert, and they compared the codes and themes. They had several discussions on the consistencies and discrepancies and finalized the themes. To measure the reliability of the theme analysis, the researcher used the interrater reliability formula that was suggested by Miles and Huberman (1994). The formula suggests that interrater reliability could be determined when the number of agreements is divided by the total number of agreements plus disagreements, and a minimum 80% of agreement is suggested for good reliability. In this study, the interrater reliability was measured as 84%. The emerging themes and categories finalized by the researcher and the expert are presented in Table 4.18. below.

Table 4.18. *Themes and Categories Regarding Alternative Assessment Applications*

<i>Themes</i>	<i>Categories</i>	<i>N</i>
Effective Learning Process	Systematic assignments/tasks	8
	Constructive feedback after tasks	7
	Ample use of the target language	5
Fairer Assessment	Benefits of rubrics	9
	Evaluation of the whole process	7
Affective Contribution	Stress-free learning environment	8
	Boosting self-confidence	5
	Higher motivation to learn	4
Suggestions for Future	Frequency of the tasks	4
	Percentage of scores in GPA	3
Comparison of Two Methods	Duration of Assessment	9
	Focus of Assessment	5

As mentioned before, the participants were also asked to keep learning/reflective journals throughout the process and answered certain questions regarding certain alternative assessment methods and how these methods affected their learning processes. As a result of the content analysis of the interviews and reflective journals, five main themes and several other categories emerged.

4.4.1. Effective Learning Process

To begin with, most of the participants believed that they experienced an effective learning process thanks to alternative assessment applications. To support this claim, eight participants stated that it was beneficial for them to complete certain tasks in a systematic manner. This issue was described by one of the participants as follows:

Our teacher assigned tasks for each week, which pushed me to get prepared for these tasks seriously. I knew that if I did not do my job well, I would receive low grades. For instance, I had presentation tasks on a regular basis, and improved my speaking dramatically. Also, I kept a portfolio and wrote a lot of paragraphs and essays. Thanks to systematic feedback sessions with my peers and teacher, I also improved my writing (Participant 2).

Another participant supported the issue in the following statement:

Our teacher assigned us tasks regularly, and I know that many students maintained their connections with the school because of this. I am a teacher as well, and I know that it is very difficult to motivate

students to attend classes regularly. We had a tight schedule of assignments this term and the students had to participate in the lessons to get good grades. As a result of this, they improved their English I think (Participant 13).

Similarly, another participant uttered the following statement on the issue:

This process kept me alert all the time since I was assigned tasks regularly and there were deadlines. I never cut off my tie with the school and neither did many other students. I think the process was challenging but considering the outcome, it was beneficial for me. (Participant 9).

Another category that emerged from the statements of seven participants regarding the systematic learning process was receiving constructive feedback after tasks. One of the participants uttered the following statement on the issue:

Receiving feedback on my mistakes, especially after writing tasks, dramatically improved my performance in writing. Sometimes my friends also helped me correct my mistakes during self-assessment sessions, but I benefited from the teacher the most (Participant 7).

The benefit of receiving constructive feedback on the learning process was mentioned by another respondent in the following statement:

During this process, I completed several tasks and received comments and feedback from the teacher. After each feedback session, I noted down my mistakes and learned from them. In the following tasks, I tried not to repeat them. In traditional exams, I do not have that chance. I take these exams at the end of the learning process without seeing my weaknesses. So, it is too late to correct them (Participant 5).

A final statement on this issue was stated by another respondent as follows:

Before this process, I had done many tasks and received scores for them, but I did not receive comprehensive feedback on my performance. During this process, on the other hand, all the tasks were assigned in a systematic way, and I received detailed feedback for my performance. In short, everything was handled in a disciplined way (Participant 13).

The last category in relation to the effective learning process was ample use of the target language. Five participants believed that they were given many opportunities to use the target language both during the preparation phase of the tasks and during their actual performances as well. One of the respondents explained this issue as in the following statement:

Our teacher assigned us various tasks during this process. We had to shoot a video, prepare a presentation for the speaking lesson on regular basis, which improved my speaking ability. We also had a debate in the last week. For these tasks, I had to get prepared very well. For example, I had to check the meaning and pronunciation of some words. I also had to rehearse my presentations at home. For the debate, I had to prepare some notes in English. Even during these preparation stages, I improved my English a lot. (Participant 10).

Another participant supported the issue with the following statement:

Giving presentations really helped me improve my speaking. I had to do this frequently, so I got fluent in English. Also, we constantly practised writing on different issues and received feedback from the teacher on our work. This improved my writing ability. In traditional exams, we cannot use the language, especially in terms of speaking, as much as we want. However, I felt that I really could speak the language during this process (Participant 6).

Lastly, one of the interviewees explained this issue as follows:

Prior to this process, learning English was all about grammar rules, reading and learning vocabulary for me. I never had a chance to use the language in a productive way. During this process, we were given a lot of tasks and used the language. For example, for the video task, we had to write a short story and we had to act out this story. During the preparation, we practised speaking a lot with my friends. Similarly, for the presentation tasks, I had to go online and make a search on my topic. After the presentations, other students also asked questions to the presenter, which promoted a lot of interactions in class (Participant, 5).

In short, participants in the experimental group believed that they experienced an effective learning process thanks to alternative assessment applications. Based on the analysis of the responses of the participants, it could be stated that the participants benefited from the regular assignment of alternative assessment tasks because they kept them alert all the time. The participants also found the grading rubrics useful because they were able to receive constructive feedback from both their peers and the teacher. Lastly, they believed that they had ample opportunities to use the language communicatively as they were engaged in several alternative assessment tasks throughout the process.

4.4.2. Fairer Assessment

The second theme that emerged from the analyses of the participants' responses was about fairer assessment. To start with, the participants believed that their performances were assessed fairly since certain rubrics were used. They also stated that they benefited from these rubrics in different ways. One of the respondents addressed the issue in the following statement:

I think the use of rubrics was useful. I was able to notice my strengths and weaknesses in this way, and I tried to improve myself in weak areas. In addition, I felt that my performance was assessed in a fair way. The rubrics provided equal assessment for all students since they included clear grading criteria and every student got what they deserved (Participant 5).

Another participant supported the issue as in the following statement:

We have never used a rubric during the previous modules. I realized I really enjoyed being assessed through rubrics. I didn't use to receive detailed feedback on my performance before. Thanks to the rubric, I was able to compensate for my weaknesses in specific areas before it is too late. For example, I realized that I had problems using correct punctuation in writing since I received low scores in that area on the rubric. I think students can learn a lot from rubrics (Participant 1).

The positive impact of rubrics on assessment is also reflected by one of the respondents as follows:

I think using rubrics in class was a good idea. Thanks to rubrics, students learned to be realistic. They received constant feedback and were confronted with their weaknesses. During the self-assessment process, they learned how to assess themselves objectively thanks to rubrics. In addition, they were roughly able to estimate their scores after they used rubrics during self-assessment. They felt they were being assessed in a systematic and fair way. I work as a teacher too, and from now on, I am planning to use rubrics to assess my students' performance assignments (Participant 13).

Another category regarding fairer assessment was the evaluation of the whole process. The participants stated that during alternative assessment applications, they had completed several tasks during the process and their performances were not assessed in a one-shot manner based on a single task. Instead, the whole learning process was considered for assessment. The following statement by one of the participants explains this issue as follows:

I strongly believe that alternative assessment applications should be used more for assessment. During this process, we were assessed for all our efforts and our teacher constantly supported us to get better. The good thing about alternative assessment it considers everything we have done in class for assessment. Some students never participate in class, but they pass the exams. I don't think they really learn the language in this way. On the other hand, some students are good at English and have actively participated in class throughout the year, but they get low scores on school exams due to some personal problems or excitement. I don't think they deserve these low scores (Participant 8).

The positive impact of evaluating the whole process was addressed by another respondent with the following statement:

I think our future should not only depend on traditional exams. These exams only assess our performance in one day and this is not a good way of assessment for me. I think there should be other options for assessment. Some students need more time to learn, but these traditional exams do not consider this and block their learning. Alternative assessment applications allow these students enough time to improve themselves as there is a long process ahead of them. I think alternative and traditional assessment tools should be used equally at school (Participant 6).

Likewise, another participant supported the issue by uttering the statement below:

As a teacher, I believe that students' performances should not be only assessed through traditional exams. When students take an exam at the end of the term, it can be too late to compensate for their weaknesses. During this process, I had to complete many tasks, especially in writing and speaking lessons. After each task, I received feedback on my mistakes, and I tried not to repeat them. By the time I took the final exam at school, I had already improved myself in many ways. If I hadn't been assessed regularly and received feedback from you, I would not have had a chance to notice my mistakes before the final exam. I think the learning process was more beneficial for me than taking some exams (Participant 13).

Based on these statements of the participants, it can be stated that the participants in the experimental group believed that alternative assessment applications provided a fairer assessment in class because the use of rubrics ensured transparency in grading. In addition, they stated that their performances were evaluated throughout the whole process in a formative manner, which enabled them to show what they can do in multiple occasions.

4.4.3. Affective Contribution

Another theme generated from the interview data was related to affective contribution. Most of the participants maintained that alternative assessment applications had a positive effect on their affective states. Three more categories emerged under affective contribution which are stress-free learning environment, boosting self-confidence, and

higher motivation to learn. One of the participants reported the following statement regarding stress-free learning environment:

I think traditional exams cause too much stress for me, so I cannot reflect my real performance on these exams. Time is limited, so I always panic and make mistakes. Normally, I perform well in class, especially in speaking and writing lessons. We completed several alternative assessment tasks this year. We both completed individual tasks and sometimes studied in groups. We had a lot of time to get prepared for these tasks, so we didn't feel stressed. I realized that I perform well with the people I know. I really enjoyed it (Participant 11).

Another participant supported the same issue in the following statement:

During the process, we were always engaged in different activities. We were assigned tasks, and we were given enough time to get prepared for these tasks. To perform well in the class, I knew that I had to put in a lot of effort, and I did. For some tasks, like video shooting tasks or debates, we studied in groups and motivated one another. We never felt stressed since it was group work, and we knew we did our best to succeed. On the other hand, in midterm or final exams, fear of failing really makes me nervous as I feel I am in a race against time. The best feature of alternative assessment is the flexibility of time for me (Participant 4).

The second category related to the affective states of learners was boosting self-confidence. The participants stated that alternative assessment applications increased their self-confidence as they completed several tasks successfully throughout the process. One of the participants supported this issue by the following statement:

I always had a lack of self-confidence in English. After Module B1, however, I started to participate in class activities since I had lots of weekly assignments. I realized I started using the language more often and this really increased my self-confidence, especially in speaking. When I failed in traditional exams, I thought I was never going to learn this language. However, when I performed badly in alternative tasks, I didn't feel like that because I could always make up for this in the following tasks (Participant 11).

Another participant explained this issue by uttering the following statement:

I am not afraid of speaking in front of people anymore. I used to be afraid of making mistakes in front of my friends, so I refrained from participating in class activities. However, giving presentations in class and participating in debates really helped me to build confidence and I started to be more active during classes (Participant 3).

The third category that emerged from the participants' responses was higher motivation. The participants believed that alternative assessment applications enhanced their motivation in class. One of the participants explained the relationship between the alternative assessment process and motivation by the following statement:

I used to feel shy and nervous during class activities. However, I completed several alternative assessment tasks during the term, and I said to myself "I can do this". I started to come to school in a more motivated way. I wish we had experienced such a learning system before. On the other hand, traditional exams usually affect students' motivation levels negatively. For instance, when a student gets a low score on a midterm exam, they tend to lose their interest in lessons (Participant 10).

Another participant mentioned the positive effect of alternative assessment on motivation in the statement below:

During alternative assessment applications, we were constantly exposed to English, and that's the reason why we came to this school. We were engaged in so many different activities that we improved our language skills in a natural way. When I realized this improvement, my prejudices against learning the language started to disappear, and my motivation increased as well (Participant 4).

Finally, one of the participants uttered the following statement on the same issue:

At the beginning of the year, the activities in our course books were interesting for me, but then I started to get bored doing the same things repeatedly. However, alternative assessment tasks were a new experience for me, and I really enjoyed them. The diversity of tasks prevented boredom in class, and I never lost my motivation in learning (Participant, 2).

Overall, it can be stated that alternative assessment applications had a positive effect on the affective states of the participants in the experimental group. The participants believed that alternative assessment methods created a stress-free environment for them, which helped them feel more self-confident in class. In addition, they participated in the classes in a more motivated way because they were engaged in diverse tasks throughout the term and used the language communicatively while performing these tasks.

4.4.4. Suggestions for Future

Although the participants mostly had positive ideas about the use of alternative assessment applications, they also offered some suggestions for improving of the process. Therefore, the fourth theme is titled as suggestions for future. This theme is divided into two categories as frequency of the tasks and percentage of scores in GPA. One of the respondents commented on the frequency of the tasks as follows:

The use of alternative assessment applications was beneficial for me. However, I also had to study for my school exams. Therefore, I felt exhausted from time to time, and I could not spend enough time on some of the tasks. In addition, we had assignments for other lessons. I think the number of tasks should be reduced a bit (Participant 1).

Another student addressed the same concern by uttering the following statement:

The tasks assigned during this process were effective for us. However, the frequency of tasks should be reconsidered. Each week, we were assigned a different task and some students, including me, felt overwhelmed. I think these tasks could be announced on a schedule at the very beginning of the term so that students could know what to expect (Participant 4).

The second category emerged as a suggestion was related to the percentage of alternative assessment scores in GPA. In accordance with the school policy, alternative assessment scores of the participants constituted only 10% of the GPA. On the other hand, the participants stated that the percentage of alternative assessment scores in GPA needed to be increased.

One of the participants explained this issue by uttering the following statement:

I think the percentage of the alternative assessment scores should be increased in calculating total scores. Some students may not reflect their real performances in midterm or final exams due to reasons such as health problems, lack of attention or time limitation. If the percentage of alternative assessment scores were higher, these students could compensate for their poor performances in traditional exams (Participant 3).

Another participant supported the issue as follows:

I think both traditional and alternative assessment methods are necessary for us. However, the percentage of alternative assessment scores need to be higher. The percentages of final and midterm exam scores, which constitute 50% and 30% of the total passing grade respectively, should be reconsidered (Participant 1).

Based on the analysis of the responses of the participants in the experimental group, it can be stated that the participants made two suggestions for the future practices of alternative assessment. They believed that the frequency of the tasks needs to be reconsidered because they felt overwhelmed due to the tight schedule of tasks. In addition, they stated that the weight distribution of alternative assessment scores should be increased in the calculation of their GPA. In other words, they believed that the ratio of traditional exams needs to be reduced in the calculation of their total scores.

4.4.5. Comparison of Two Methods

The last theme generated from the responses of the participants was related to the differences between alternative and traditional assessment methods. This theme is divided into two more categories as duration of assessment and focus of assessment. Most of the participants agreed that there is a difference between the two assessment methods in terms of the duration of the assessment. They stated that alternative assessment encompasses a long process while traditional assessment occurs within a limited time. One of the participants uttered the following statement on this issue:

In traditional assessment methods, the exam content is limited, and there is time limitation as well. How are we supposed to reflect our general proficiency in just two hours? During the alternative assessment process, we completed several tasks with different skills, and we had enough time to get prepared. We received regular feedback after the tasks, so we performed better in the following tasks. (Participant 12).

Another respondent supported the issue with the following statement:

During alternative assessment applications, I improved myself in English, especially in terms of speaking and writing. The more I produced something and received feedback on my performance, the better I got in these skills, which led to permanent learning for me. In traditional assessment, they assess our speaking performance in only five minutes. Some students who normally perform well in class might get too nervous, or they might be asked about unfamiliar topics. When there is a time limitation, they could also have problems in the writing section, which can cause inaccurate assessment for them (Participant 4).

An interesting statement by one of the participants on the issue is illustrated below:

If my goal was just to pass the prep class, I would support the use of traditional methods since we take some exams, and we can pass the class easily. However, I am at this school to learn the language, therefore I support the use of alternative assessment methods (Participant 6).

Likewise, another participant uttered an interesting statement on the time issue as follows:

I support the use of traditional assessment methods since these methods include exams that last a short time, and if we study hard for them, we pass easily. Alternative assessment methods are time consuming and tiring. Each student has a different life outside school, and some students may not cope with alternative assessment tasks (Participant 7).

The second category emerging from the responses of participants was the difference in the focus of assessment. The participants believed that alternative assessment applications focus on how learners produce the language while traditional assessment methods focus on checking whether learners have acquired the necessary knowledge or information provided to them during the lessons. One of the participants explained this issue in the following statement:

Traditional exams force learners to memorize theoretical information such as grammar rules. Students usually cannot use this information in their real lives. On the other hand, we were assigned several tasks during alternative assessment applications, and we used the language in a productive way. I think this is what we need in real life (Participant 15).

Another respondent uttered the following statement on the same issue as in the following:

During the alternative assessment process, we were assigned diverse tasks with different skills. We searched for information and got prepared for these tasks and practised a lot. During this preparation stage, learning took place. After we performed our tasks, we received scores, which was motivating for us. We felt we did something practical. Traditional exams are only useful for checking our theoretical knowledge, and this is also necessary, I think. However, we feel stressed because these exams decide whether we pass or fail (Participant 13).

Finally, one of the participants approached the issue by the following statement:

I still keep the assignments that I completed during the alternative assessment process since I produced something in English during these assignments. They are important to me. For traditional exams, I try to memorize some rules and vocabulary, but I forget most of them after the exams. I don't feel I learned them permanently (Participant 11).

Overall, the participants in the experimental group emphasized two main differences between alternative assessment and traditional assessment. They believed that alternative assessment evaluates the whole learning process in a formative manner whereas traditional exams only assess them at the end of the term within a limited time, which causes stress and anxiety in learning. In addition, they stated that alternative assessment focuses more on improving productive skills of learners while traditional exams only test what they know and remember about a specific lesson.

CHAPTER 5: DISCUSSION, CONCLUSION, AND SUGGESTIONS

This study aims at finding out the effect of applying alternative assessment tools on a group of Turkish EFL students' academic achievement and investigating the attitudes of these students towards learning English before and after the implementation of alternative assessment tools in a preparatory school. To achieve this aim, the current study attempted to answer three main research questions as well as dealing with a number of sub-problems related to these questions. The following section presents the research questions and their sub-problems in relation to the findings of the previously conducted studies in the field.

5.1. Discussion

5.1.1. Discussion of the First Research Question and Its Sub-problems:

Research question 1 aimed to find out whether there was a statistically significant difference between the overall alternative assessment scores and traditional assessment scores of the participants in the experimental group. To find out this, the overall alternative assessment scores, and traditional assessment scores of the participants in the experimental group were compared. The comparison of the participants' overall alternative and traditional assessment scores revealed that there was a statistically significant difference between the overall alternative assessment scores ($M=83.46$, $SD=10.19$) and the traditional assessment scores ($M=72.34$, $SD=6.06$) of the participants in the experimental group $t(31)=5.33$, $p=.00$. In other words, the participants in the experimental group performed much better during alternative assessment applications and obtained significantly higher scores as a result of their performances.

The reason for this difference could stem from the fact that the alternative assessment scores of the participants in the experimental group and their traditional exam assessment scores were obtained in different ways. As for the alternative assessment scores, their writing scores were determined based on their portfolio performances. As explained in the methodology part in detail, the participants completed several paragraph and essay tasks on various topics that were covered in their coursebooks. Throughout the process, their performances were assessed through certain checklists and rubrics in a formative manner. In addition, the participants performed self and peer assessments during the process to monitor their own progress and receive supportive feedback from their peers. Most importantly, they received corrective feedback from their teacher during one-to-one conferences. In other

words, the participants were informed about their strengths and weaknesses on certain issues during these systematic feedback sessions and improved their writing skills (Cauley & McMillan, 2010). The participants' speaking scores, on the other hand, were determined based on their speaking performances throughout the process. The participants were engaged in presentation tasks, debates and video shooting activities which were designed in parallel with their coursebooks and the course syllabus. To determine their speaking scores, their performances were assessed through certain rubrics. Similar to the writing process, their speaking performances were also assessed in a formative manner. Since reading and listening are regarded as covert and receptive skills, the participants' reading and listening scores had to be determined based on their performances in integrated tasks, namely speaking and writing tasks during alternative assessment applications, and their performances were assessed through certain rubrics in a formative manner.

On the other hand, the experimental group participants' traditional scores were determined based on their performances in the final exam. As explained in the methodology part, the participants took a final exam at the end of the module in a summative manner. This exam consists of four sections including listening, writing, reading, and speaking skills, and the participants had to complete this exam in just two hours. The listening and reading sections of the final exam included items such as multiple-choice and gap-filling questions which are regarded as traditional test items (Belle, 1999, as cited in Wikström, 2008; Brown & Hudson, 1998). In the listening section, the participants were allowed to listen to the tracks only twice and answer the related questions. Similarly, in the reading section, they had to read two reading texts and answer the following questions within a limited time. On the other hand, in the writing section of the final exam, the participants had to complete an essay writing task for the writing section of the final exam. Although this task seems similar to their alternative assessment tasks, the participants had a limited time to write their essays, which might have affected their performances and motivation to learn in a negative way (Shohamy, 1982).

As one can understand from the explanations above, it is not surprising that the experimental group participants attained higher alternative assessment scores compared to their traditional assessment scores. To summarize the main reasons for these findings, it could be stated that the alternative assessment scores of the participants were gathered in a formative manner while their traditional assessment scores were obtained in a summative manner. In other words, they were constantly assessed through alternative assessment applications, received constructive feedback on their performances and their alternative

assessment scores were calculated at the end of the process. In contrast, their traditional scores were gathered through one-shot traditional tests at the end of the term within a limited time. Therefore, the participants might not have reflected their real performances in these tests (Barootchi & Keshavarz, 2002). Most importantly, the participants could have suffered from exam anxiety during traditional assessment practices. On the other hand, time was flexible during alternative assessment practices, so the participants had ample time to get prepared for their tasks without feeling anxious or stressed (Berry, 2008; Dikli, 2003; Korkmaz, 2006). All in all, all these factors could have led to the difference between the participants' alternative assessment and traditional assessment scores.

Secondly, to answer *research question 1a*, the alternative assessment and traditional assessment scores of the participants in the experimental group were compared in terms of writing skills. The statistical analyses revealed that there was not a statistically significant difference between the alternative assessment scores ($M= 88.46, SD=8.81$) and the traditional assessment scores ($M=85.46, SD=9.78$) of the participants in the experimental group in terms of writing skills $t(31)=1.50, p=.14$. This result suggests that the participants attained similar scores in their alternative and traditional assessment tasks in terms of writing skills. The reason for this similarity could stem from the fact that the participants were often engaged in integrated paragraph and essay writing tasks which were prepared in parallel with the school's writing syllabus. In addition, as explained in the methodology part of this dissertation, they performed self and peer assessments and received constructive feedback from their teacher during one-to-one conferences and also from their peers. They also kept all their written works in their portfolios, which gave them a chance to monitor their progress in writing. Most importantly, they experienced a stress-free writing experience since there was no time limitation and they were able to receive constructive feedback both from their peers and teacher through the rubrics, which helped them improve their writing skills dramatically (Hamp-Lyons & Condon, 2000). In this sense, the implementation of portfolios could be regarded as a constructive learning method in this process (Buyukduman & Sirin, 2010). All in all, it could be stated that the participants completed so many writing tasks and improved their writing skills during the alternative assessment process that they also performed well in their traditional writing exams. This finding, in parallel with the findings of several previously conducted studies, suggests that alternative assessment tasks, when applied and assessed systematically, could serve as an appropriate method to assess learners' writing skills (Aydın & Başöz, 2010; Farahian & Avarzamani, 2018; Khodadady & Khodabakhshzade, 2012; Lam & Lee, 2010; Obeiah & Bataineh, 2016).

The next sub-problem was addressed in *research question 1b*, which aimed to compare the alternative assessment scores and the traditional assessment scores of the participants in the experimental group in terms of reading skills. The statistical analyses revealed that there is a statistically significant difference between the alternative assessment scores ($M=80.93$, $SD= 18.02$) and the traditional assessment scores ($M=72.50$, $SD= 13.50$) of the participants in the experimental group in terms of reading skills $t(31)=2.20$, $p=.03$. This finding was, in fact, expected for several reasons. Firstly, as mentioned in the discussion part of *research question 1*, the participants' alternative assessment scores in reading were attained through integrated tasks, namely speaking, and writing tasks since reading is a covert and receptive skill. In other words, the participants were engaged in communicative tasks based on the reading texts in their coursebooks either in written or spoken form. In this way, they had a chance to delve into the reading texts and use the content for communicative purposes. In other words, they were required to move from reading to either speaking or writing. On the other hand, their traditional reading scores were obtained from their standardized multiple-choice reading tests which were applied in a summative manner. Secondly, the participants had ample time to get prepared for their integrated tasks during the alternative assessment applications. Therefore, they did not suffer from time pressure and may not have felt stressed. On the other hand, they had to answer the traditional reading tests within a certain amount of time, which might have affected their performances negatively. In addition, the reading texts might have been unfamiliar to them, making the exam process even harder. All in all, all these factors might have led to the difference between the participants' alternative and traditional assessment scores in terms of reading skills. In the literature, it has been claimed that the use of reading content for integrative tasks could yield some positive results. According to Willis (1981), learners tend to become more interactive in reading classes thanks to integrated tasks since they are pushed to produce the language rather than only getting prepared for the exams. In addition, their self-confidence might be boosted as they show their reading skills through communicative tasks. Shohamy (1998) also supports this issue stating that integrating receptive skills into productive skills through communicative tasks will lead to more effective and permanent teaching and learning for meaningful and constructive purposes.

To answer *research question 1c*, the alternative assessment and traditional assessment scores of the participants in the experimental group were compared in terms of listening skills. The statistical analyses revealed that there was a significant difference between the alternative assessment scores ($M=80,37$, $SD=15,48$) and traditional assessment

scores ($M=59.56$, $SD=15.16$) of the participants in the experimental group in terms of listening skills $t(31)=5.26$, $p=.00$. Like reading skill, listening skill is also regarded as a covert and receptive skill. Therefore, the participants' listening scores were attained through integrated tasks, namely speaking, and writing tasks, during alternative assessment applications in a formative manner. On the other hand, their traditional listening scores were obtained from their standardized multiple-choice reading tests which were applied in a summative manner. Considering the fact that the participants were only allowed to listen to the tracks twice during their traditional listening tests and had to demonstrate their micro and macro skills to answer the questions, it is not surprising that they attained lower scores in these tests compared to their alternative assessment listening scores, which were gathered through integrated tasks in a formative manner. Since alternative assessment methods aim to capture the overall array of skills and abilities through integrated tasks, learners need to be able to perform tasks that require the integration of different dimensions of the language (Tierney, Carter, & Desai, 1991). Therefore, the participants in the experimental group were supposed to integrate the listening content in their coursebooks into certain communicative tasks designed either in spoken or written form.

The last sub-problem was addressed in *research question 1d*, which attempted to compare the alternative assessment and traditional assessment scores of the participants in the experimental group in terms of speaking skills. At the end of the statistical analyses, it was found that there was not a statistically significant difference between the alternative assessment scores ($M=87.40$, $SD= 12.32$) and the traditional assessment scores ($M=86.06$, $SD= 11.31$) of the participants in the experimental group in terms of speaking skills $t(31)=.45$, $p=.65$. During the alternative assessment applications, the participants were engaged in several speaking tasks such as giving a presentation, participating in a debate, and shooting a video. In addition, they were also involved in question-answer sessions after the presentations and debates. Therefore, there was a lot of interaction in the classroom, and they had ample opportunities to use the language communicatively. Their performances in these speaking tasks were assessed through certain rubrics in a formative manner. Their traditional speaking scores, on the other hand, were determined based on their performances in the final speaking exam, which lasted approximately five to ten minutes. The fact that their alternative assessment speaking scores and traditional assessment speaking scores were similar was expected. As the participants were engaged in so many communicative tasks during the alternative assessment applications, they improved their speaking skills dramatically and performed well in their final speaking exam as well.

This finding suggests that learners' speaking performances could be assessed in a formative manner through alternative assessment tasks. If alternative assessment tasks are applied and assessed systematically through appropriate tools, they could replace the traditional speaking exams, which would yield several positive results. To begin with, learners could have a chance to use the language through meaningful tasks, collaborate with one another during these tasks, and thus participate in the lessons more (Abu Rahmah & Al Humaidi, 2012; Brown & Hudson, 1998; Inbar-Lourie & Donitsa-Schmidt, 2009; Poehner & van Compernelle, 2011). Secondly, they could perform these tasks without feeling stressed since they have enough time to prepare for these tasks (Cirit, 2015; Dikli, 2003). Since most learners suffer from anxiety during speaking activities, they may not reflect their real competencies in traditional speaking tests, which are usually applied in a summative manner within a short time. Students only receive a grade for a snapshot of their performances in these traditional tests which may not be a fair assessment. On the other hand, alternative assessment tasks could be a better option to assess speaking skill since students' speaking performances were not assessed based on a single exam. Instead, they are assessed for their overall performances during the whole process in a formative manner (Alderson & Banerjee, 2001), which also has a positive washback effect on their learning (Bayram, 2015). During this process, they would receive constant feedback on their performances and work on their weaknesses for the following tasks (Areiza-Restrepo, 2013). Therefore, they could reach their ultimate levels of performances. Lastly, the use of rubrics would ensure transparency in grading, which would build trust between the teacher and students. When students know about the assessment criteria, they tend to believe that their performances are assessed fairly. In addition, rubrics serve as a diagnostic tool through which learners can detect their strengths and weaknesses, which would guide and aid learners in their learning processes (Knoch, 2009; Weigle, 2005).

5.1.2. Discussion of the Second Research Question and Its Sub-problems:

The aim of *research question 2* was to find out whether there was a statistically significant difference between the overall traditional assessment scores of the participants in the experimental group and the control group. First of all, the overall traditional scores of both groups were compared, and the statistical analyses revealed that there was a significant difference between the overall traditional assessment scores ($M= 72.34$, $SD=6.06$) of the participants in the experimental group and the traditional assessment scores ($M=65.85$, $SD=7.93$) of the participants in the control group $t(64)=-3.71$, $p= .00$. This finding suggests

that the participants in the experimental group outperformed the participants in the control group in the final exam. In other words, alternative assessment applications had a positive effect on the overall academic achievement of the participants in the experimental group.

The fact that alternative assessment applications had a positive impact on the academic achievement of students in the current study was supported by a number of previously conducted studies. In one of the earliest studies, Kavaliauskiene, Kaminskienė, & Anusienė (2007) attempted to explore the effect of alternative assessment on the linguistic development of 96 Lithuanian EFL learners in the context of English for Specific Purposes (ESP). Their study also aimed at finding out the challenges of implementing alternative assessment in the institution. Portfolios were used as a means of alternative assessment in the study, and the participants kept their vocabulary tasks, PowerPoint Presentations, and essays in their portfolios. At the end of the study, the statistical analyses revealed that, echoing the findings of the present study, alternative assessment contributed to the linguistic developments of the participants to a great extent. In addition, the participants' reflections on the usefulness of different tasks guided teachers to overcome the challenges of alternative assessment in the institution.

Another study on the effectiveness of alternative assessment was conducted by Yurdabakan and Erdogan (2009). In their experimental study, portfolio was implemented as an alternative assessment method in the experimental group while the control group experienced a conventional teaching method. At the end of the study, the scores of the participants in each group were compared in terms of writing, reading, and listening skills. The findings revealed that the participants in the experimental group achieved significantly higher scores than the participants in the control group in writing, though no such significance was observed in reading or listening skills. The researchers of the study believe that, supporting the findings of the present study, portfolio is an appropriate assessment method since it contributes to linguistic development of learners, increases their responsibilities, and motivates them in class.

In a more recent study, Burnaz (2011) conducted a case study with 21 participants at a state university in Türkiye to find out whether portfolios provide any benefits to the participants as claimed in the literature. In addition, the study tried to investigate the perceptions of the participants regarding portfolio keeping as an alternative assessment method. The qualitative findings of the study revealed that the participants were observed to improve their language skills, particularly writing and vocabulary skills. In addition, they claimed that portfolio keeping led to permanent learning for them. Echoing the findings of

the present study, Fajarsari (2016) conducted a study to examine the perceptions of 40 Indonesian EFL learners towards self-assessment, peer assessment and performance assessment as alternative assessment methods as well as exploring the benefits of such methods. In addition, the study aimed to find out these EFL learners' favorite alternative assessment methods. At the end of the statistical analyses, it was found that the majority of the participants supported use of alternative assessment methods together with traditional ones in EFL classes. The participants' responses in the survey indicated that alternative assessment methods enable them to foster the use of four language skills in an integrated way, which was also supported by the findings of the present study. In addition, the participants believed that alternative assessment methods created an interactive and cooperative learning environment, and thus, increased their motivation in learning the target language. Regarding their favorite alternative assessment method, the participants favored self-assessment more than peer assessment or performance assessment though all of these assessment methods provided several benefits for them. They maintained that self-assessment provided them an opportunity to check their products and helped them increase their grades in class. Moreover, they stated that, as supported by the findings of the present study, performance assessment enabled them to use the language in productive tasks and improved their productive skills. Finally, they believed that peer assessment created a cooperative learning environment in the classroom and also reduced the stress of being assessed by the teacher.

In her MA thesis, Özuslu (2018) tried to investigate a group of EFL learners' perceptions regarding the use of performance-based assessment tasks (PTs), as a supplementary component of the assessment system in a preparatory school. In addition, the study examined the expectations and views of the instructors and administrators on the implementation of PTs. The study had a mixed method research design, and the participants consisted of 126 students, 60 instructors and two administrators. During the treatment process, the participating students completed several PTs for their grammar, reading/vocabulary, and listening/speaking lessons. The findings of the study revealed that all the participants had a positive attitude towards the use of PTs in terms of planning, application, scoring, learning outcomes and program consistency phases. As for the impact of PTs on language proficiency, echoing the findings of the present study, the participants believed that PTs greatly enhanced their lexical knowledge, improved their reading skills, and contributed to their general proficiency in English. In addition, although majority of the stakeholders recommended some revisions for the PTs, they all agree that they should be a part of the assessment system in the preparatory school in the following years. Based on

these findings, the researcher of the study claims that PTs need to be considered as a supplement to traditional assessment methods at preparatory schools.

In her doctoral dissertation, Shakoori (2022) attempted to explore the forms of alternative assessment applications in college-level in EFL context. The study also tried to explore the perceptions and experiences of the participants concerning alternative assessment applications as well as examining how their language learning is affected by such assessment. This case study included 11 participants who were taking English courses at a state university in Saudi Arabia. The data were collected through open-ended questionnaires, semi-structured interviews, and focus group interviews. The participants were engaged in several alternative assessment techniques and methods such as presentations, projects, discussions, peer assessments, self-assessments, and portfolios. At the end of the study, the statistical analyses revealed that alternative assessment methods substantially contributed to the participants' language proficiency, which supports the findings of the present study, as well. All of the participants claimed to have observed improvement in their language abilities. They believed that portfolios greatly helped them to improve their writing skills whereas discussions and presentations enabled them to advance their speaking skills. Since they had to listen to their peers and teacher during discussions, the participants also improved their listening skills. In addition, they had a chance to improve their spelling, punctuation and improved their grammar during self and peer assessment sessions. Lastly, the participants asserted that alternative assessment tasks developed their research skills since they had to search for information to get prepared for the presentation and discussion tasks. In short, the participants mentioned several benefits and merits of alternative assessment methods that dramatically contributed to their language learning and proficiency, which are in parallel with the findings of the present study as well. In addition to linguistic benefits, the use of alternative assessment methods yielded some positive results in terms of affective states of the participants. Since the participants experienced an interactive learning process, they were constantly supported, and praised by their peers and the teacher through constructive feedback. As a result, their anxiety level decreased while their motivation and self-confidence boosted in class, which is another finding of the present study.

As mentioned above, many studies in the literature, along with the current study, supported the use of alternative assessment applications in language classrooms. Based on the results of these studies, it can be stated that the use of alternative assessment techniques and methods in language classrooms could yield positive results in terms of language improvement of learners. Since these methods allow them to be engaged in various

communicative and meaningful tasks, students' motivation and self-confidence also increase in class. Therefore, such methods need to be incorporated into the syllabus and assessment procedures of language institutions. As well as investigating the effect of alternative assessment methods on general language proficiency of learners, the present study aimed to reveal the effects of these methods on the development of four language skills.

To answer *research question 2a*, the traditional assessment scores of the participants in the experimental group and the control group were compared in terms of writing skills. The statistical analyses showed that there was not a significant difference between the traditional assessment scores ($M= 85.46$, $SD=9.78$) of the participants in the experimental group and the traditional assessment scores ($M=83.17$, $SD=10.61$) of the participants in the control group $t(64)=-.91$, $p= .36$ in terms of writing skills. This finding was indeed expected since the participants in both groups were engaged in similar writing activities during the process. They followed the same writing syllabus and used the same coursebook. In addition, they received constructive feedback from their teachers on their written works. As a result, they improved their writing skills by the end of the process and performed well in their traditional final exam as well. Although there was not a statistically significant difference in their traditional writing scores, the participants in the experimental group mentioned several benefits regarding the alternative assessment methods used in their writing lessons. The qualitative data also revealed that the participants enjoyed having a portfolio for their writing classes since it enabled a systematic writing process with regular feedback sessions. In addition, the participants claimed that having self and peer assessment sessions and the use of rubrics helped them improve their writings a lot. One of the participants uttered the following statement on the value of feedback:

Receiving feedback on my mistakes, especially after writing tasks, dramatically improved my performance in writing. Sometimes my friends also helped me to correct my mistakes during peer assessment sessions, but I benefited from the teacher feedback most (Participant 7).

Another participant mentioned the benefit of rubrics as in the following:

We have never used a rubric during the previous modules. I realized I really enjoyed being assessed through rubrics. I didn't use to receive detailed feedback on my performance before. Thanks to rubric, I was able to compensate for my weaknesses in specific areas before it is too late. For example, I realized that I had problems in using correct punctuation in writing since I received low scores in that area on the rubric. I think students can learn a lot from rubrics (Participant 1).

Several studies in the literature also found that alternative assessment methods are effective in improving the writing skills of learners as well as increasing their motivation in writing classes. As for portfolio assessment, the findings of several studies revealed that

learners benefited from keeping a portfolio in EFL writing classes in many ways. Portfolios not only improved learners' writing skills but also improved their motivation in L2 writing classes (Arslan & Gümüş, 2020; Burnaz, 2011; Demirel & Duman, 2015; Farahian & Avarzamani, 2018; Fathi et al., 2020; Obeiah & Bataineh, 2016; Yurdabakan & Erdogan, 2009). Likewise, the positive impact of self- and peer assessment and one-to-one conferences was also supported by several studies (Fathi et al., 2021; Irajii et al., 2016; Javaherbakhsh, 2010; Kızıllı, 2019; Medfouni, 2014). These studies found that self- and peer assessment are effective in improving learners' writing skills since they create a collaborative learning environment and provide continuous feedback from different sources in class. In addition, learners gain new perspectives through meaningful interactions with their peers (Cheng & Warren, 2005).

To answer *research question 2b*, the traditional assessment scores of the participants in the experimental group and the control group were compared in terms of reading skills. The statistical analyses showed that there was a significant difference between the traditional assessment scores ($M=75.50$, $SD=13.50$) of the participants in the experimental group and the traditional assessment scores ($M=64.82$, $SD=14.63$) of the participants in the control group $t(64)=-.221$, $p=.03$ in terms of reading skills. This finding suggests that the participants in the experimental group outperformed the participants in the experimental group in the final exam in terms of reading skill. This could be explained by the fact that the participants in the experimental group had to do more than only answering the traditional comprehension questions on their reading coursebooks. As explained in the methodology section in detail, the participants were engaged in certain integrated tasks that were prepared based on their reading coursebooks. Since reading is a covert and receptive skill, their performances in reading had to be assessed through productive tasks. For instance, they had to write a paragraph, or an essay based on a reading text they had read. To do so, they had to delve into the texts so that they could incorporate the content into their written works, which improved their micro and macro skills in reading as well. In addition, they were supposed to give presentations on the issues they had covered in reading lessons. They again had to read the texts carefully to prepare their presentations. As one can see, the participants in the experimental group did not just answer the comprehension questions related to the texts on their coursebooks but also performed several communicative tasks, which made them read the texts several times and helped them to improve their reading skills. Another important benefit here is that traditional question items such as multiple-choice or gap filling exercises only assess the lower-level cognitive skills of learners while integrated tasks can

assess learners' higher-level cognitive skills as well (Alhareth & Dighrir, 2014; Al Sadaawi, 2010). Therefore, the participants in the experimental group had a chance to improve their integrated competencies during the process.

As the next step, the *research question 2c* aimed at comparing the traditional assessment scores of the participants in the experimental group and the control group in terms of listening skills. At the end of the statistical analyses, it was found that there was a significant difference between the traditional assessment scores ($M= 59.56$, $SD=15.16$) of the participants in the experimental group and the traditional assessment scores ($M=45.52$, $SD=16.36$) of the participants in the control group $t(64)=-3.60$, $p= .00$ in terms of listening skills. This finding suggests that the participants in the experimental group obtained considerably higher scores than the participants in the control group in the final exam in terms of listening skill. One possible explanation for this difference could be that the participants in the experimental group were engaged in several different communicative tasks such as presentations, debates, and video shooting tasks. During these activities, they had to listen to their peers and teacher carefully so that they could respond appropriately. In addition, as in reading skill, they also had to complete some integrated tasks based on the conversations and lectures that they had listened to. In order to incorporate the content of the lectures into their written or oral tasks, they had to do listen to the tracks several times. In short, the participants were constantly exposed to the target language and received a great amount of input from their peers, the teacher, and lesson materials, which could have contributed to their listening skills. Brooks and Brooks (2001) focused on this issue and stated that learners needed to be encouraged to interact with their peers and teachers as much as possible to gain meaningful learning experiences. Although the participants in the experimental group had better scores, their mean scores seemed not so high. This could stem from the fact that they were only able to listen to the tracks twice in the final exam and answer the multiple-choice questions in a short period of time. Therefore, their performances might have been affected negatively due to stress and time pressure. To avoid this negative outcome, their listening skills could be assessed through alternative assessment methods in a formative manner.

Finally, to answer *research question 2d*, the traditional assessment scores of the participants in the experimental group and the control group were compared in terms of speaking skills. The statistical analyses revealed that there was a significant difference between the traditional assessment scores ($M= 86.06$, $SD=11.31$) of the participants in the experimental group and the traditional assessment scores ($M=78.82$, $SD=12.49$) of the

participants in the control group $t(64)=-2.46, p=.01$ in terms of speaking skills. This finding suggests that the participants in the experimental group performed significantly better than the participants in the control group in the final exam in terms of speaking skills. As mentioned in the methodology part in detail, the participants in the experimental group performed several oral tasks throughout the process. For instance, they gave presentations on the issues covered in their coursebooks, participated in debates, and tried to defend their ideas on certain issues against their peers. They also did some video shooting activities in groups. All these activities appeared to have pushed them to use the language for communicative purposes. In addition, the preparation of these tasks also contributed to their speaking skills since they had to rehearse their presentations and acting for the video shooting tasks several times prior to their actual performances. This preparation stage not only improved their speaking abilities but also reduced their stress levels since they had enough time to get ready for their actual performances. Lastly, they had ample opportunities to practice the language during question-answer sessions after presentations and debates, which triggered a lot of meaningful interaction in class and motivated them to participate in the lesson more. As one can see, the participants in the experimental group were able to improve their speaking skills through diverse tasks and felt more motivated and relaxed during the performance of these tasks.

In the literature, several studies supported the findings of the present study. To illustrate, in her thesis, Pesen (2016) conducted a case study to find out the impact of Drama and Music Portfolio (DMP) on the speaking skills of 34 Turkish EFL learners. The quantitative analyses revealed that there was a significant increase in the participants' speaking scores after the implementation of DMP. The qualitative analyses also revealed that a great majority of the participants held positive attitudes towards the use of DMP and preferred to be assessed through such methods rather than traditional speaking exams. Safari and Koosha (2016) also tried to investigate the effects of speaking portfolios on 64 Iranian EFL learners' speaking ability. The study also aimed to explore the perceptions of these learners' regarding the use of speaking portfolios in their speaking classes. The treatment process lasted 12 sessions and for each session, the participants in the experimental group were assigned individual, pair or group speaking tasks. All of these performances, reflection papers and assessments were stored in the participants' speaking portfolios. The statistical analyses revealed that participants in the experimental groups achieved significantly higher scores than their counterparts in the control groups. In addition, the participants in the experimental groups were observed to develop positive attitudes towards the use of speaking

portfolios. They also claimed that they improved their speaking skills dramatically and felt more confident in speaking lessons as time passed.

The qualitative findings of the present study also revealed the participants in the experimental group found alternative assessment tasks beneficial and believed that they had improved their speaking skills and felt more comfortable during their speaking performances. One of the participants uttered the following statement on the positive impact of the alternative assessment on her speaking skills:

Our teacher assigned us various tasks during this process. We had to shoot a video and prepare a presentation for the speaking lesson on a regular basis, which improved my speaking ability. We also had a debate in the last week. For these tasks, I had to get prepared very well. For example, I had to check the meaning and pronunciation of some words. I also had to rehearse my presentations at home. For the debate, I had to prepare some notes in English. Even during these preparation stages, I improved my English a lot (Participant 10).

Another respondent supported the issue with the following statement:

I always had a lack of self-confidence in English. After Module B1, however, I started to participate in class activities since I had lots of weekly assignments. I realized I started using the language more often and this really increased my self-confidence, especially in speaking. When I failed in traditional exams, I thought I was never going to learn this language. However, when I performed badly in alternative tasks, I didn't feel like that because I could always make up for this in the following tasks (Participant 11).

As can be understood from the findings above, alternative assessment methods had a positive impact both on learners' speaking skills and their affective states. Since learners are involved in meaningful interactions through authentic real-life tasks, they are observed to make progress in speaking and feel more motivated in speaking classes (Reeves, 2000, as cited in Nasab, 2015).

5.1.3. Discussion of the Third Research Question and Its Sub-problems:

Research question 3 aimed to explore the attitudes of the participants in the experimental group towards learning English through a scale. To start with, as *research question 3a* indicates, the pretest and posttest scores of the participants in the control group were compared in terms of their attitudes towards learning English. The statistical analyses revealed that there was a statistically significant difference between the pretest scores ($M=107.39$, $SD=9.07$) and posttest scores ($M=97.00$, $SD=11.51$) of the participants in the control group $t(37)=4.86$, $p=.00$ in terms of their attitudes towards learning English. This finding suggests that the participants in the control group had a lower level of attitude towards learning English at the end of the process though they were observed to have a relatively high attitude before the study. One explanation for this finding could be that although the

participants in the control group followed the same syllabi and used the same coursebooks, their performances were assessed through traditional methods. As explained in the literature review part, there are several negative features of traditional assessment methods. First, they are unable to assess the real competencies of learners through meaningful tasks (Poehner & van Compernelle, 2011; Sidek, 2012). These methods force learners to memorize information and rules to pass certain exams rather than making them use the language communicatively (Burnaz, 2011). Second, they are usually applied at the end of the learning process in a summative manner within a limited time. As a result, students generally suffer from stress and anxiety during these exams, and this could have a negative effect on learners' motivation in learning English (Falchikov, 2005; Knight, 2001). Most importantly, traditional exams mostly focus on assessing receptive skills, ignoring particularly speaking skills (Burnaz, 2011).

Next, the *research question 3b* attempted to compare the pretest and posttest scores of the participants in the experimental group in terms of their attitudes towards learning English. At the end of the analyses, it was found that there was not a statistically significant difference between the pretest scores ($M=105.86, SD=9.15$) and posttest scores ($M=102.37, SD=7.71$) of the participants in the experimental group $t(36)=1.82, p=.07$ in terms of their attitudes towards learning English. This finding suggests that the participants in the experimental group still maintained their high attitudes toward learning English at the end of the process in spite of a slight decrease in their posttest scores. As explained in the methodology section in detail, the participants in the experimental group followed the same syllabi and used the same coursebooks like the participants in the control group. However, their performances were assessed through a number of alternative assessment methods in a formative manner throughout the process. Considering the results of *research question 3b*, it could be inferred that alternative assessment methods had a positive impact on the learning process of the participants in the experimental group, which helped them to maintain their high attitudes towards learning English as well.

The qualitative findings of the present study also revealed that the use of alternative assessment applications had a positive effect on the perceptions and attitudes of the participants towards learning English in the experimental group. One of the participants uttered the following statement on the positive impact of the alternative assessment on motivation:

I used to feel shy and nervous during class activities. However, I completed several alternative assessment tasks during the term, and I said to myself "I can do this". I started to come to school in

a more motivated way. I wish we had experienced such a learning system before. On the other hand, traditional exams usually affect students' motivation levels negatively. For instance, when a student gets a low score on a midterm exam, they tend to lose their interest in lessons (Participant 10).

Another respondent supported the same issue by reporting the following statement:

During alternative assessment applications, we were constantly exposed to English, and that's the reason we are at this school. We were engaged in so many different activities that we improved ourselves in language skills in a natural way. When I realized this improvement, my prejudices against learning the language started to disappear, and my motivation increased as well (Participant 4).

In addition to motivation, another participant explained the positive effect of alternative assessment methods on their self-confidence by the following statement:

I always had a lack of self-confidence in English. After Module B1, however, I started to participate in class activities since I had lots of weekly assignments. I realized I started using the language more often and this really increased my self-confidence, especially in speaking. When I failed in traditional exams, I thought I was never going to learn this language. (Participant 11).

Another participant explained this issue by uttering the following statement:

I am not afraid of speaking in front of people anymore. I used to be afraid of making mistakes and thus remained silent in class. Giving presentations in class and participating in debates really helped me build confidence and I started to be more active during classes (Participant 3).

Several studies also found that alternative assessment methods affected learners' perceptions and attitudes towards learning positively. For instance, Yurdabakan and Erdogan (2009) conducted a study and found that portfolio keeping motivated learners in learning English and increased their sense of responsibility as well. In a similar study, Burnaz (2011) found that portfolio keeping increased the motivation of learners and made them autonomous in learning. In her MA study, Fajarsari (2016) investigated a group of EFL learners' perceptions towards alternative assessment methods. The results of the study revealed that alternative methods not only helped the participants to obtain high grades at school but also motivated them to learn the target language. Ardianti and Mauludin (2017) also conducted a qualitative study and found that authentic assessment motivated EFL learners to have a reading routine, which enabled them to expand their vocabulary and get familiar with English discourse. Additionally, Suwaed (2018) conducted a mixed-method study with a group of Libyan EFL students regarding their perceptions towards portfolios as an alternative assessment method. The findings of the study indicated that portfolio keeping enhanced the students' motivation in learning, especially in terms of writing skills. Besides, in her MA thesis, Özuslu (2018) conducted a study on the use of performance tasks as an alternative assessment method at a preparatory school in Türkiye. It was found that most of the participating instructors expressed their positive views on the use of performance tasks since they seem to affect students' affective states positively, particularly in terms of

motivation, learner autonomy and self-confidence. In short, all of these studies, echoing the findings of the present study, indicate that alternative assessment applications have positive effects on students' affective states in foreign language classrooms.

5.2. Conclusion, Suggestions and Pedagogical Implications

Since traditional assessment methods are unable to serve the purposes of communicative language teaching practices, there has been an increasing interest in new assessment methods (Herman et al., 1992). Barootchi and Keshavarz (2002) stated that traditional assessment methods must be supported by certain alternative assessment methods that consider not only the product but also the process of learning. Therefore, it has been widely acknowledged that new or alternative assessment methods are needed to assess students' language skills through meaningful tasks and gather evidence regarding their progress and motivation in language learning (Brown & Hudson, 1998; Stiggins, 2005; Sidek, 2012). Therefore, the aim of this study is to find out the effect of applying alternative assessment tools on a group of EFL students' academic achievement and investigate the attitudes of these students towards learning English before and after the implementation of alternative assessment tools in a preparatory school.

The participants of this quasi-experimental study consisted of 75 preparatory class students whose ages ranged from 18 to 21. There were 38 participants in the experimental group, while the control group had 37 participants. Both experimental and control group participants were placed in B1 classes based on the placement test scores. The treatment process lasted 16 weeks in two modules as B1 and B1+. At the beginning of the treatment process, an attitude scale towards learning English was administered to both experimental and control group participants to gather data about their attitudes towards learning English before the treatment. During the treatment process, the participants in the experimental group participants were engaged in several alternative assessment tasks such as presentations, debates, video shootings and role-plays. Their performances on these tasks were assessed through certain checklists and rubrics in a formative manner. They also kept a portfolio, performed self and peer-assessment, and had one-to-one conference sessions to receive continuous support and feedback from their peers and the teacher during their writing lessons. The main purpose of implementing alternative assessment methods was to encourage the participants to use the language communicatively both inside and outside the classroom. In addition, the researcher aimed to provide an opportunity for the participants to

integrate four language skills through meaningful tasks. As mentioned above, the performances of the experimental group participants were assessed by means of certain rubrics, checklists, and one-to-one conferences with the teacher in a formative manner. Thanks to these tools, they received constructive feedback and got scores on their performances systematically. At the end of the process, they had a collection of alternative assessment scores for each skill.

As for the participants in the control group, they used the same coursebooks and followed the same syllabi as the participants in the control group. For each skill lesson, they completed the existing exercises in their textbooks and did the worksheets assigned as homework. For the reading and listening lessons, they completed multiple choice and gap-filling exercises in their coursebooks. For the speaking lesson, they had some communicative activities such as group discussions and pair work activities on the unit topics and also prepared some presentations on some of them. Their performances during the speaking lessons were assessed in a holistic manner by their teachers. For the writing lessons, they wrote several paragraphs and essays on certain issues and received corrective feedback from their teachers on their performances. Unlike the participants in the experimental group, they did not have to keep a portfolio for their writing lesson. They also did not perform self or peer assessment during their writing practices. During the process, the participants were also assigned many worksheets and took several quizzes for all skills so that they could be prepared for their exams. As one can see, the participants in the control group were also engaged in various activities during the process like the participants in the experimental group. However, their performances were only assessed through traditional assessment methods, which included multiple choice and true-false items as well as gap-filling exercises, in a summative manner.

At the end of the treatment process, the same attitude scale towards learning English was administered to both experimental and control group participants to see whether there was a statistically significant difference between their pretest and posttest scores based on the scale data. In addition, alternative assessment scores of the experimental group participants were gathered for each skill. To increase the reliability of these scores, an external rater, who works in the School of Foreign Languages and holds a PhD degree in the ELT field, also assessed the data throughout the process. In addition to the alternative assessment scores, traditional exam scores of both experimental and control group participants were obtained from the testing office of the School of Foreign Languages. Finally, semi-structured interviews were conducted with 15 experimental group participants

to gather their views and feelings on the implementation and effectiveness of alternative assessment applications. After the data collection process ended, both quantitative and qualitative statistical analyses were conducted to answer the related research questions.

The statistical analyses revealed that there was a statistically significant difference between the overall alternative assessment scores and the traditional assessment scores of the participants in the experimental group. In other words, the participants in the experimental group performed much better during alternative assessment applications and obtained significantly higher scores as a result of their performances. The reason for this difference could stem from the fact that the experimental group participants' alternative assessment scores were obtained through integrated tasks in a formative manner with no time pressure while their traditional exam assessment scores were obtained through traditional tests in a summative manner. On the other hand, it was found that there was not a statistically significant difference between the alternative assessment scores and traditional assessment scores of the participants in the experimental group in terms of writing skill. This could be explained by the fact that the participants completed so many writing tasks and received constructive feedback and support from their peers and the teacher in a systematic way. Consequently, their writing skills gradually improved during the alternative assessment process, and they performed well in their traditional writing exams as well. This finding, in parallel with the findings of several previously conducted studies, suggests that alternative assessment tasks, when applied and assessed in a systematic way, could serve as an appropriate method to assess learners' writing skills (Aydın & Başöz, 2010; Farahian & Avarzamani, 2018; Khodadady & Khodabakhshzade, 2012; Lam & Lee, 2010; Obeiah & Bataineh, 2016). Therefore, the negative effects of traditional writing tests could be eliminated through the use of alternative assessment methods in writing classes. It was also found that there was a significant difference between the alternative assessment and traditional assessment scores of the participants in the experimental group in terms of reading and listening skills. The mean scores indicated that the participants performed better during the alternative assessment tasks than they did in their traditional tests. This finding was, in fact, expected since reading and listening are receptive skills and the participants' performances were assessed through integrated tasks (i.e., speaking and writing tasks) in a formative manner with no time pressure for these skills. However, their traditional tests, which included multiple choice and gap filling test items, were applied in a summative manner. It is therefore not surprising that the participants attained higher scores from alternative assessment tasks. As learners are expected to use the language in an integrated

way in their everyday lives, alternative assessment tasks could be used to encourage them to use the language in this way. Finally, there was not a significant difference between the alternative assessment and traditional assessment scores of the participants in the experimental group in terms of speaking skill. This could be explained by the fact that the participants were engaged in so many communicative tasks during the alternative assessment applications that they improved their speaking skills dramatically and performed well in their final speaking exam as well. This finding suggests that alternative assessment methods could be a better alternative to traditional speaking tests for several benefits. First, learners are given many opportunities to use the language communicatively. Second, they do not suffer from time pressure and stress, which could affect their performance negatively. Most importantly, they are assessed in a formative manner rather than being assessed for a snapshot of their performances (Brown & Hudson, 1998; Dikli, 2003; Inbar-Lourie & Donitsa-Schmidt, 2009; Poehner & van Compernelle, 2011).

The statistical analyses also revealed that there was a significant difference between the overall traditional assessment scores of the participants in the experimental group and the overall traditional assessment scores of the participants in the control group. The mean scores indicated that the participants in the experimental group attained much higher scores than those in the control group. This finding suggests that alternative assessment applications had a positive impact on the language learning process and academic achievement of the participants in the experimental group. As for language skills, the participants in the experimental group obtained significantly higher scores than the participants in the control group in terms of speaking, reading, and listening skills. However, no such significant difference was observed between the scores of the two groups in terms of writing skill despite a slight difference in mean scores in favor of the experimental group. Considering these findings, it could be suggested that alternative assessment methods contribute to the improvement of four language skills of learners when applied in a systematically way.

Finally, the participants' attitudes towards learning English were investigated for both groups. The statistical analyses indicated that the participants in both groups had high levels of positive attitudes towards learning English prior to the intervention. At the end of the intervention, however, the statistical analyses revealed that the participants in the experimental group were observed to maintain their high attitudes towards learning English whereas the participants in the control group had lower attitudes. This suggests that alternative assessment methods had a positive impact on the experimental group participants' attitudes towards learning English. The qualitative data gathered from the

participants in the experimental group also supported this finding. Since alternative assessment methods include diverse meaningful and communicative tasks, the participants were observed to have a high level of motivation and self-confidence during classes. In addition, the formative nature of assessment eliminated time pressure and reduced the stress levels of the participants during the process. Therefore, it could be suggested that alternative assessment methods could be incorporated into classrooms to boost the motivation levels of learners and facilitate their learning.

5.2.1. Pedagogical Implications

The findings of this dissertation suggest a number of implications for several stakeholders in foreign language education. First of all, the findings of the study suggest that alternative assessment applications have a positive impact on the language learning process and academic achievement of EFL learners. It is a well-known fact that learners suffer from the intensity of traditional exams, which put a lot of pressure on them. What is worse, they fail to reflect their grammatical and lexical knowledge in a communicative way on these exams. Therefore, teachers need to incorporate integrated and communicative alternative assessment methods along with traditional assessment methods to improve the learning conditions in their classes. Secondly, alternative assessment methods also affect learners' affective states positively. The qualitative findings of the study suggest that since alternative assessment methods consist of diverse authentic and communicative tasks which are aligned with the learning outcomes, learners tend to feel more motivated in class and participate in lessons more. Therefore, a student-centered learning atmosphere could be created. Another affective benefit of alternative assessment methods is learners feel more confident using the language, particularly in speaking and writing classes. Since learners complete several productive tasks and receive continuous support and feedback during the process, they gradually improve their language abilities and gain self-confidence in using the language. In addition, it is suggested that school administrators and other stakeholders need to consider including more alternative assessment practices in their assessment policies. Based on the qualitative findings of the current study, it could be stated that learners expect to be engaged in alternative assessment methods and assessed through certain rubrics and checklists since they believe that this is necessary for a fair and accurate assessment of language learning. However, it is of great importance that school directors provide teachers with the necessary training on how to implement alternative assessment applications in an appropriate and systematic way. To help this process, the present study offers an alternative assessment

implementation guideline whose details are explained below.

5.2.1.1. Suggested guideline for alternative assessment implementation. The present study suggests an alternative assessment implementation guideline by identifying and explaining the steps followed. Considering the overall positive impact of alternative assessment applications, it is of great importance to propose a guideline based on the treatment process and suggestions of the participating students regarding the whole process. The guideline consists of three parts: the *preparation stage*, the *implementation stage*, and the *evaluation stage*.

In the preparation stage, the stakeholders such as heads of departments, testing, and syllabus coordinators could come together and identify the objectives of alternative assessment for each level considering the outcomes of CEFR. The objectives could range from enhancing the language skills of learners through integrated tasks to improving their critical thinking skills. Then, some focus group discussions could be organized to determine the scope of alternative assessment. In other words, the skills that will be included in the alternative assessment tasks and how they could be integrated need to be determined. In addition, the administrators need to decide how alternative assessment procedures will be utilized in the evaluation criteria of their institutions.

In the next step, the stakeholders could start designing some alternative assessment tasks based on the topics in the selected coursebooks. It is of great importance that these tasks should also match the learning objectives of each skill course for validity issues. Furthermore, the designed tasks need to be authentic and meaningful to real life contexts as well as catering to different learning styles and intelligences. While designing the tasks, the stakeholders could use the content in reading and listening coursebooks, and design some integrated productive tasks based on these. In this way, the receptive skills could be incorporated into productive skills by students in an integrated way. Most importantly, the stakeholders need to decide on certain checklists and rubrics for the assessment of students' task performances so that they can establish clear and measurable assessment criteria to evaluate students' performances. They have three options here. First, they can adopt an already existing rubric or checklist if it is suitable for their purposes. Second, they can adapt these existing rubrics and checklists with minor revisions, which was the case in this study. Finally, they can design a new rubric or checklist from scratch for their purposes (Perlman, 2003). For speaking skill, the stakeholders could select the most appropriate alternative assessment techniques. Performance-based tasks such as presentations, role-plays, debates,

interviews, and video shooting tasks are commonly employed techniques for this skill. For writing skill, they could consider a portfolio as an alternative method. They could also include self- and peer assessment practices during writing classes. Since listening and reading are receptive skills, they could also be assessed through writing and speaking tasks in an integrated manner. Project-based tasks could be employed here as they create an opportunity for students to integrate various skills to accomplish a task.

Since teachers will carry out all these procedures, they obviously have a crucial role in the implementation and evaluation of alternative assessment practices. As teachers mostly employ summative assessment tools to evaluate their students' progress in language learning, they may lack in knowledge and expertise in alternative assessment applications. Therefore, enhancing, and broadening teachers' knowledge base regarding alternative assessment applications sounds critical. To make their teachers more familiar with alternative assessment procedures such as designing appropriate tasks and reliable scoring rubrics, school administrators need to provide adequate in-house training for teachers before the implementation stage starts.

In the implementation stage, the appropriateness of the tasks and rubrics and checklists could be piloted in certain classes at different levels. Before this piloting process, students need to be informed about the alternative assessment process. For instance, they need to know what type of alternative assessment techniques and methods will be employed during the process. In addition, assessment tools such as checklists and rubrics should be introduced to them as transparency of grading is an essential feature of alternative assessment. In this way, students will be informed what is expected from them. During the piloting process, both teachers and students could keep a journal to reflect on the process. After a certain amount of time, maybe after one module or semester, the views and feelings of the students and teachers regarding the process could be gathered through some surveys or interviews. Based on the findings, the tasks and rubrics/ checklists could be revised by the stakeholders for future practices since reviewing and updating practices is another key component of alternative assessment.

In the evaluation stage, the views and feelings of the students and teachers could be gathered through a final survey and reflection papers with open-ended questions. In this way, the effectiveness of the process could be evaluated in a detailed way, and the necessary revisions and preparations could be planned for future practices.

On the other hand, it is important to note that each education context has its own characteristics and conditions. Therefore, the guideline explained in detail above could be

expanded or adapted to different language education contexts. In spite of the demanding nature of alternative assessment applications, the process could be successfully handled if the necessary training and guidance were provided to both teachers and learners. Table 5.1. illustrates the main steps of the suggested alternative assessment implementation guideline.

Tablo 5.1. *Suggested Alternative Assessment Implementation Guideline*

Preparation Stage:

- Identify the objectives of alternative assessment with all stakeholders.
- Determine the scope of alternative assessment and its place in general evaluation criteria.
- Decide on alternative assessment methods for each skill or for the integration of various skills based on the objectives.
- Design alternative assessment tasks for each skill or the integration of skills based on the objectives.
- Establish clear grading criteria by selecting checklists and rubrics for the assessment of the tasks.
- Provide adequate in-house training on alternative assessment applications for teachers before the implementation stage starts.

Implementation Stage:

- Inform teachers and students about the alternative assessment process.
- Introduce tasks and assessment criteria such as rubrics and checklists.
- Conduct the piloting of the tasks and assessment tools in certain classes at different levels.
- Receive feedback from students and teachers on the effectiveness of the process.

Evaluation Stage:

- Gather the final views and feelings of both teachers and students through surveys and interviews.
 - Discuss the benefits and challenges of the alternative assessment process.
 - Make the necessary revisions and preparations for future practices.
-

5.2.2. Suggestions for Further Research

The current study has offered important insights into the nature and value of alternative assessment. However, it also has certain limitations, which could pave the way for further studies regarding alternative assessment applications. To start with, the present study was conducted in only one school, therefore its findings cannot be generalized to other institutions. A more comprehensive study including different settings and participants could

yield more generalizable results. In addition, in the current study, the application, storing, and assessment of the participants' performances were exhausting for both the participants and the teacher. Therefore, another study could employ some technological tools to facilitate the application, storing, assessment of alternative assessment tasks and explore the effects of using such tools on the effectiveness of the process. Finally, the current study did not aim to reveal the perceptions of the instructors regarding alternative assessment applications. Therefore, a further study that includes several instructors is needed so that the perceptions of instructors regarding alternative assessment applications can be explored in detail.

REFERENCES

- Abbas, Z. (2012). Difficulties in using methods of alternative assessment in teaching from Iraqi Instructors' point of view. *Al-Fatih Journal*, 48(6), 23-45.
- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do? In D. Corrigan, J. Dillon, & R. Gunstone (Eds.). *The professional knowledge base of science teaching* (pp. 205–221). The Netherlands: Springer.
- Abu Rahmah, M. I., & Al Humaidi, S. H. (2012). Enhancing ELT through self-assessment. *ELT Research Journal*, 1(1), 31-52.
- Afshar, H. S. & Bastami, T. (2012). The effect of portfolio assessment on Iranian pre intermediate EFL learners' autonomy. *TELL*, 6(1),71-102. doi:10.22132/TEL.2012.54921
- Agustina, R. (2011). Portfolio as alternative assessment in language testing, *Eksplanasi*, 6(1), 1-6.
- Airasian, P. & Russell, M. (2008). *Classroom assessment: Concepts and applications*, (7th edition.). McGraw-Hill Higher Education, Columbus, OH.
- Akerman, R. & I. Neale. (2011). *Debating the evidence: An international review of current situation and perceptions*. Research report, CfBT Education Trust. ISBN 978-1-907496-55-4.
- Al Alhareth, Y., & Al Dighrir, I. (2014). The assessment process of pupils' learning in Saudi education system: A literature review'. *American Journal of Educational Research*, 2(10), 883-891. doi:10.12691/education-2-10-6
- Alderson, J., & Banerjee, J. (2001). Language testing and assessment. *Language Teaching*, 34, 213 – 236. doi: 10.1017/S0261444802001751
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency. The interface between learning and assessment*. London, UK: Continuum.
- Al Sadaawi, A. S. (2010). Saudi national assessment of educational progress (SNAEP). *International Journal of Education Policy and Leadership*, 5(11), 1-14.
- Altiner, C. (2018). Preparatory school students' English language learning motivation: a sample from Turkey. *Universal Journal of Educational Research*, 6(8), 1729-1737. doi: 10.13189/ujer.2018.060814
- Anderson, J. O., & Bachor, D. G. (1998). A Canadian perspective on portfolio use in student assessment. *Assessment in education: Principles, Policy & Practice*, 5(3), 353-379.
- Anderson, L. W. (1999). Rethinking Bloom`s taxonomy: Implications for testing and assessment. Retrieved from ERIC database. (ED435630).

- Ardianti, T. M., & Mauludin, L. A. (2017). Students' responses on the application of authentic assessment in EFL reading class. *The Journal of English Language and Literature*, 1(2), 110-122. doi: 10.31002/METATHESIS.V1I2.470
- Areiza-Restrepo, H. N. (2013). Role of systematic formative assessment on students' views of their learning. *Profile Issues in Teachers' Development*, 15(2), 165-183.
- Arslan, R. Ş., & Gümüş, S. N. (2020). The effect of portfolio-keeping on EFL young learners' writing achievement and their motivation towards writing. *Journal of Narrative and Language Studies*, 8(14), 130–150.
- Aydın, S., Başöz, T. (2010). The attitudes of pre-service teachers towards EFL writing. *Journal of Language and Linguistic Studies* 6(2), 54-69.
- Aydin, S. (2010). EFL writer's perceptions of portfolio keeping. *Assessing Writing*, 15(3), 194-203. doi: 10.1016/j.asw.2010.08.001
- Bachman, L. F. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Backman, L. O. & Klinghammer, S. J. (2006). *Shaping the way we teach English instructor's manual*. Washington: University of Oregon.
- Bahous, R. (2008). The self-assessed portfolio: A case study. *Assessment & Evaluation in Higher Education*, 33(4), 381-393. doi:10.1080/02602930701562866
- Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Heinle & Heinle, Boston.
- Baker, L. (2010). *What probably works in alternative assessment*. (CRESST Report 772). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Reprinted from paper originally presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, April, 1991.
- Baleghizadeh, S., & Zarghami, Z. (2012). The impact of conferencing assessment on EFL students' grammar learning. *Profile Issues in Teachers' Professional Development*, 14(2), 131-144.
- Balliro, L. (1993). What kind of alternative? Examining alternative assessment. *TESOL Quarterly*, 27(3), 558-561.
- Barootchi, N., & Keshavarz, M. H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, 44 (3), 279–288. doi:10.1080/00131880210135313
- Bayram, E. (2015). *An investigation of language assessment preferences of pre- service EFL teachers and underlying factors*. Unpublished master thesis. Karadeniz Technical University, Trabzon.

- Belle, D. (1999). *Traditional assessment versus alternative assessment*. Unpublished master thesis. Kean University of New Jersey, ED431012. Retrieved from <http://edres.org/eric/ED431012.htm> (2007-04-08).
- Benzehaf, B. (2017). Exploring teachers' assessment practices and skills, *International Journal of Assessment Tools in Education*, 4 (1)1-18. doi:10.21449/ijate.254581
- Berg, B. L. (2007). *Qualitative research methods for the social sciences*. London: Pearson.
- Berry, R. (2008). *Assessment for learning*. Hong Kong University Press.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals*. Handbook I: Cognitive domain. New York: Longman.
- Bogdan, R.C., & Biklen, S.K. (2007). *Qualitative research for education: An introduction to theory and methods*, 5th ed. Boston: Allyn & Bacon.
- Braun, H. I., & Mislavy, R. (2005). Intuitive test theory. *Phi Delta Kappan*, 86 (7), 489-497. doi: <https://doi.org/10.1177/003172170508600>
- Brawley, N. (2009). *Authentic assessment versus traditional assessment: A comparative study*. Unpublished master thesis. Coastal Carolina University. Retrieved from <https://digitalcommons.coastal.edu/honors-theses/146>.
- Broadfoot, P. (2003). Performance assessment in perspective: International trends and current English experience. In A. Craft (Ed.), *Primary education assessing and planning learning* (pp. 36-65). Taylor & Francis e-Library, London.
- Brooks, M., & Brooks, J. (2001). *In search of understanding: The case for constructivist classrooms*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL quarterly*, 32(4), 653-675. doi:<https://doi.org/10.2307/3587999>
- Brown, H. D. (2000). *Principles of language learning and teaching* (4th Ed.). Englewood Cliffs, NJ: Prentice Hall.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman.
- Brown, J. D. (2005) *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw-Hill.
- Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy*. White Plains, NY: Longman.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practice*. Pearson Longman.
- Bryman, A. (2016). *Social research methods* (5th Ed.). London Oxford University Press.

- Burnaz, E. Y. (2011). *Perceptions of EFL learners towards portfolios as a method of alternative assessment: A case study at a Turkish state university*. Unpublished master thesis. Middle East Technical University, Ankara.
- Butler, S. M., & Mcmunn, N. D. (2006). *Understanding and using assessment to improve student learning*. PB Printing: United States of America.
- Buyukduman, I., & Sirin, S. (2010). Telling ELT tales out of school, learning portfolio (LP) to enhance constructivism and student autonomy. *Procedia Social and Behavioral Sciences*, 3, 55-61. doi:10.1016/j.sbspro.2010.07.012
- Büyüköztürk, Ş. (2013). *Sosyal Bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum*. Ankara: Pegem Akademi Yayıncılık.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Cauley, K. M., & McMillan, J. H. (2010). Formative Assessment Techniques to Support Student Motivation and Achievement. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, 83(1), 1–6. doi:10.1080/00098650903267784
- Chang, C., & Tseng, K. (2009). Use and performances of Web-based portfolio assessment. *British Journal of Educational Technology*, 40(2), 358-370. doi:358-370. 10.1111/j.1467-8535.2008.00885.x
- Charvade, M. R., Jahandar, S., & Khodabandehlou, M. (2012). The impact of portfolio assessment on EFL learners' reading comprehension ability. *Canadian Center of Science and Education*, 5(7), 129-139. doi:10.5539/elt.v5n7p129
- Chen, Y., & Hoshower, L.B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation, *Assessment & Evaluation in Higher Education*, 28(1), 71-88. doi: 10.1080/02602930301683
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360–389. doi:10.1191/0265532204lt288oa
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22 (1), 93–121. doi: <https://doi.org/10.1191/0265532205lt298oa>
- Chirimbu, S. (2013). Using alternative assessment methods in foreign language teaching. Case study: Alternative assessment of business English for university students. *Scientific Bulletin of the Politehnica University of Timișoara Transactions on Modern Languages*, 12, 91-98. Retrieved from <https://dspace.upt.ro/xmlui/handle/123456789/5882>.
- Cirit, N. C. (2014). *Perceptions of ELT pre-service teachers toward alternative assessment via Web 2.0 tools: A case study at a Turkish state university*. Unpublished master thesis. Middle East Technical University, Ankara.

- Coombe, C., & Barlow, L. (2004). The reflective portfolio: Two case studies from the United Arab Emirates, *English Teaching Forum* 41(1), 26-35.
- Coombe, C., Folse, N., & Hubley, N. (2007). *A practical guide to assessing English language learners*. USA: The University of Michigan Press.
- Coombe, C., & Hubley, N. (2011). Fundamentals of language assessment manual: *PASAA Assessment Tools in Education*, 48 (1) 58-78, Thailand. doi:10.3126/nelta.v12i1.3425
- Cornett, C. E. (1999). *Whole language, whole learning*. Phil Delte Kappa Educational Foundation.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209 –240). Thousand Oaks, CA: Sage
- Creswell, J.W. & Plano-Clark, V.L. (2006). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd Ed.). Thousand Oaks, CA, US: Sage Publications, Inc.
- Creswell, J.W. (2009). *Research design: Qualitative, quantitative, and mixed methods and approaches*. Thousand Oaks, California: SAGE Publications.
- Cresswell, J.W. (2012). *Educational research: Planning, conducting and evaluating qualitative and quantitative research* (4th Ed.). Boston: Pearson Education Inc.
- Crick, R. D., & Yu, G. (2008). Assessing learning dispositions: Is the Effective lifelong learning inventory valid and reliable as a measurement tool? *Educational Research*, 50(4), 387-402. doi: 10.1080/00131880802499886.
- Cunningham, G. K. (1998). *Assessment in the classroom*: London. The Falmer Press.
- Çakir, C. (2013). Standard assessment and alternative assessment in English language teaching program. *Gazi University Journal of Gazi Educational Faculty* 33 (3), 531-548.
- Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. London: Routledge/Falmer.
- Davies, P. (2006). Peer assessment: Judging the quality of students work by comments rather than marks. *Innovations in Education and Teaching International*, 43(1), 69-82. doi:10.1080/14703290500467566

- Davison, C., & Leung, C. (2009). Teacher-based assessment. *TESOL Quarterly, Special Issue*, 43(3), 389–562. doi:10.1002/j.1545-7249.2009.tb00242.x
- Demir, M. (2021). Alternative assessment methods in primary education: Review and future directions. In S. A. Kiray & E. Tomevska-Ilievska (Eds.), *Current studies in educational disciplines* (pp. 227–288). ISRES Publishing, Turkey.
- Demirel, M., & Duman, H. (2015). The use of portfolio in English language teaching and its effects on achievement and attitude. *Procedia - Social and Behavioural Sciences*, 191, 2634–2640. <https://doi.org/10.1016/j.sbspro.2015.04.598>.
- Denzin, N. K. (1978). *The research act: A theoretical orientation to sociological methods* (2nd Ed.). New York: McGraw-Hill.
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. Macmillan Publishing, New York.
- Dietel, R. J., Herman, J. L., & Knuth, R. A. (1991). *What does research say about assessment?* NCREL, Oak Brook.
- Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *The Turkish Online Journal of Educational Technology*, 2(3), 13-19.
- DiRanna, K., Osmundson, E., Topps, J., Barakos, L., Gearhart, M., Cerwin, K., Carnahan, D. & Strang, C. (2008). *Assessment-centered teaching: A reflective practice*. Thousand Oaks, CA: Corwin Press.
- Dochy, F. (2001). A new era: Different needs, new challenges. *Research Dialogue in Learning and Instruction* 2 (1), 11-20. doi:10.1016/S0959-4752(00)00022-0
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Driscoll, M. P. (1994). *Psychology of learning for instruction*. Allyn & Bacon. Boston.
- Driscoll, M. (2000). *Psychology of learning for instruction*. Allyn & Bacon. Boston.
- Duban, N., & Küçükylmaz, E. A. (2008). Sınıf öğretmeni adaylarının alternatif ölçme değerlendirme yöntem ve tekniklerinin uygulama okullarında kullanımına ilişkin görüşleri. [Primary school student teachers' opinions regarding the use of alternative assessment methods and techniques in practice schools]. *Ilkogretim Online*, 7(3), 769- 784.
- Duff, P. A. (2014). Communicative language teaching. In M. Celce-Murcia, D. Brinton, & M.A. Snow (Eds.) *Teaching English as a second or foreign language*, 4th edition,(15-30). Boston, MA: National Geographic Learning/Heinle Cengage Learning.
- Erbay Cetinkaya, Ş. (2017). Rethinking portfolio-based assessment in writing courses: An investigation of students' perspective, *Journal of Narrative and Language Studies*, 5(9), 41-56.

- Estrin, E. T. (1993). *Alternative assessment: Issues in language, culture, and equity, knowledge brief*. San Francisco, CA: WestEd. ERIC ED 373 082.
- Fajarsari, L.A. (2016). *Students' perceptions to alternative assessment in English learning at SMA Kristen Satya Wacana Salatiga*. Unpublished master thesis. Indonesia. Retrieved from https://repository.uksw.edu/bitstream/123456789/9405/2/T1_112010124_Full%20textpdf.
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: Routledge Farmer.
- Farahian, M., & Avarzamani, F. (2018). The impact of portfolio on EFL learners' metacognition and writing performance. *Cogent Education*, 5(1), 1–21. doi: <https://doi.org/10.1080/2331186X.2018.1450918>.
- Fasold, R. W., & Connor-Linton, J. (2006). *An introduction to language and linguistics*. Cambridge: Cambridge University Press.
- Fathi, J., Derakhshan, A., & Safdari, M. (2020). The impact of portfolio-based writing instruction on writing performance and anxiety of EFL students. *Polish Psychological Bulletin*, 51(3), 226–235. doi: 10.24425/ppb.2020.134729
- Fathi, J., Afzali, M., & Parsa, K. (2021). Self-assessment and peer-assessment in EFL Context: An Investigation of Writing Performance and Writing Self-efficacy. *Critical Literary Studies*, 3(1) 211-232.
- Fenner, S. D. (2013). *Implementing the CCSS for English learners: The changing role of the ESL teacher*. TESOL/NCATE Press, Virginia.
- Firth, A., & Wagner, J. (2007). Second/Foreign language learning as a social accomplishment: Elaborations on a reconceptualized SLA. *The Modern Language Journal*, 91, 800-819. doi: <http://dx.doi.org/10.1111/j.1540-4781.2007.00670.x>.
- Fox, J. (2017). Using portfolios for assessment/alternative assessment. In Shohamy E., Or I., May S. (Eds.), *Language testing and assessment. Encyclopedia of Language and Education*, pp. 135-148. Auckland: Springer.
- Franklin, J. (2002). *Assessing assessment: Are alternative methods making the grade?* *Curriculum Update*, Retrieved from www.ascd.org.
- Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics*, 18(3), 402-423. doi: <https://doi.org/10.4219/jaa-2007-495>
- Fuchs, L. S. (1995). *Connecting performance assessment to instruction: A comparison of behavioral assessment, mastery learning, curriculum-based measurement, and performance assessment*. ERIC Digest E530. (ERIC Document Reproduction Service No. ED 381 984).
- Geberew, M., T. (2014). EFL classroom assessment: Teachers' practice and teaching techniques adjustment in Ethiopia, *Academic Journal of Education Research, and Review*, 9(20) 1071- 1089. doi: <https://doi.org/10.5897/ERR2014.1857>

- Gelbal, S., & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. [Teachers' perception of competence about the measurement and evaluation methods and problems they encounter] *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Ghazizadeh, M., & Bazargani, D. (2019). *Alternative assessment: The impact of self-assessment vs. peer-assessment on Iranian intermediate EFL learners' paragraph writing ability*. 3rd National Conference on new trends in English language teaching and testing. Tehran.
- Ghoorchaei, B., & Tavakoli, M. (2020). Students' perceptions about writing portfolios: A case of Iranian EFL students. *RELJ* 8(1), 21-42. doi: 10.30486/RELJ.2019.669076
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Routledge.
- Gonzales, G. C. (2003). Assessment in English language classrooms: Constructing and empowering meaningful learning. *Revista Pensamiento Actual*, 4(5), 84-92.
- Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Thousand Oaks: Corwin Press.
- Grabin, L. A. (2007). *Alternative assessment in the teaching of English as a foreign language in Israel*. Unpublished doctoral dissertation. University of South Africa, South Africa.
- Greenstein, L. (2010). *What teachers really need to know about formative assessment*. Alexandria, Va.: ASCD.
- Gubrium, J. F., & Holstein, J. A. (2002). *Handbook of interview research: Context and method*. Thousand Oaks, CA: Sage Publications.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership* 60 (5), 6-11.
- Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics*, 15, 212-226.
- Hamp-Lyons, L. (1992). Holistic writing assessment for LEP students. *Second National Research Symposium on Limited English Proficient Student Studies: Focus on Evaluation and Measurement*. Washington.
- Hamp-Lyons, L. (1995). *Portfolios with ESL writers: What the research shows*. Paper presented at the annual convention of Teachers of English as a Second Language, Long Beach, CA.
- Hamp-Lyons, L. (1997). Washback, impact, and validity: Ethical concerns. *Language Testing*, 14(3), 295–303. doi:10.1177/026553229701400306

- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research*. Cresskill: Hampton Press.
- Hamp-Lyons, L., & Condon, W. C. (2000). *Portfolios and college writing*. Cresskill, Nj: Hampton.
- Hancock, C. R. (1994). *Alternative assessment and second language study: What and why?* (Report No. ED376695 1994-07-00). Washington: Center for Applied Linguistics (CAL). (ERIC Clearing house on Languages and Linguistics Washington DC).
- Hargreaves, A. & Fullan, M. (1998). *What's worth fighting for out there?* New York: Teachers' College Press.
- Harp, B. (1991). *Assessment and evaluation in whole language programs*. Norwood, MA: Christopher-Gordon Publishers.
- Harwell, M. R. (2011). Research design: Qualitative, quantitative, and mixed methods: Pursuing ideas as the keystone of exemplary inquiry. In C. Conrad, & R. C. Serlin (Eds.), *The Sage handbook for research in education: Pursuing ideas as the keystone of exemplary inquiry*, (pp. 147-164), (Second Edition ed.). Sage.
- Hatipoğlu, Ç. (2017). History of English language teacher training and English language testing and evaluation (ELTE) education in Turkey. In Yasemin Bayyurt and Nicos S. Sifakis (Eds), *English Language Education Policies and Practices in the Mediterranean Countries and Beyond* (pp. 227-257). Frankfurt: Peter Lang.
- Herdiawan, D. R. (2018). Communicative competences in EFL learning. *Journal of English Language Learning (JELL)*, 2 (1), 43-46.
- Herman, J. L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. (Alexandria, VA, Association for Supervision and Curriculum Development).
- Higgins, D. (2011). Why reflect? Recognising the link between learning and reflection. *Reflective Practice*, 12, 583-584. doi: <http://dx.doi.org/10.1080/14623943.2011.606693>
- Hill, B. C., & Ruptic, C. (1994). *Practical aspects of authentic assessment: Putting the pieces together*. Norwood, MA: Christopher-Gordon Publishers.
- Hirvela, A., & Pierson, H. (2000). Portfolios: Vehicles for authentic self-assessment. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 13-48). Mahwah, NJ: Lawrence Erlbaum Associates.
- Huang, S. C. (2016). Understanding learners' self-Assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education*, 41, 803-820. doi: <https://doi.org/10.1080/02602938.2015.1042426>.
- Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5, 8-11. doi: 10.1017/CBO9780511667190.048

- Huerta-Macias, A. (2002). Alternative assessment: Responses to commonly asked questions. In Richards, Jack C., & Renandya, Willy A. (Eds). *Methodology in language teaching* (pp 338-343). USA: Cambridge University Press.
- Hughes, A. (1990). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. doi: 10.1177/0265532208090158
- Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, 16(2), 185-204. doi:10.1080/09695940903075958
- Iraji, H. R., Enayat, M. J., & Momeni, M. (2016). The effects of self- and peer-assessment on Iranian EFL learners' argumentative writing performance. *Theory and Practice in Language Studies*, 6(4), 716–722. doi:10.17507/tpls.0604.08
- Irons, A. (2008). *Enhancing learning through formative assessment and feedback*. Routledge: U.S.A and Canada.
- Jacobs, G. M., & Farrell, T. S. (2001). Paradigm shift: Understanding and implementing change in second language education. *TESL-EJ*, 5(1), 1-18. Retrieved from: <http://www-writing.berkeley.edu/tesl-ej/ej17/a1.html>.
- Jacobsen, D., Eggen, P. & Kauchak, D. (2002). *Methods for teaching: Promoting student learning*. NJ: Merrill Prentice Hall.
- Janisch, C., Liu, X., & Akrofi, A. (2007). Implementing alternative assessment: Opportunities and obstacles. *The Educational Forum*, 71(3):221-230. doi: 10.1080/00131720709335007
- Javaherbakhsh, M. R. (2010). The impact of self-assessment on Iranian EFL learners' writing skill. *English Language Teaching*, 3(2), 213-218. doi:10.5539/elt.v3n2p213
- Johannessen, O., & Redecker, C. (2013). Changing assessment towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79-96. doi: 10.2307/23357047
- Kalra, R., Sundrarajun, C., & Komintarachat, H. (2017). Using portfolio as an alternative assessment tool to enhance Thai EFL students' writing skill. *Arab World English Journal*, 8 (4), 292-302. doi: <https://dx.doi.org/10.24093/awej/vol8no4.20>.
- Karim, A., Amir, A., Darman, M., & Muhammad, Y. (2018). The effectivity of authentic assessment-based character education evaluation model. *TEM Journal* 7 (3), 495–500. doi:10.18421/TEM73-04

- Kastrati, S. (2013). Student self-evaluation in Albanian language. *Journal of Educational and Social Research MCSER* 3 (7), 431-435. doi:10.5901/jesr.2013.v3n7p431
- Kaufman, D. (2004). Constructivist issues in language learning and teaching. *Annual Review of Applied Linguistics*, 24, 303-319. doi:10.1017/S0267190504000121.
- Kavaliauskienė, G., Kaminskienė, L., & Anusienė, L. (2007). Reflective practice: Assessment of assignments in English for specific purposes. *Journal of the European Association of Languages for Specific Purposes*, 14, 149–166.
- Kenneth, O.G. (2008). *Teaching through role playing*. Retrieved from <https://bible.org/seriespage/3-teaching-through-role-playing>.
- Khodadady, E., & Khodabakhshzade, H. (2012). The effect of portfolio and self -assessment on writing ability and autonomy. *Journal of Language Teaching and Research*, 3(3), 518-524. doi: <https://doi.org/10.4304/jltr.3.3.518-524>.
- Kırkgöz, Y. (2005). English language teaching in Turkey: Challenges for the 21st century. In G. Braine (Ed.), *Teaching English to the world: History, curriculum, and practice*. (pp. 159-175). Mahwah: Lawrence Erlbaum Associates.
- Kırkgöz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations. *Regional Language Centre Journal*, 38(2), 216-228. doi: 10.1177/0033688207079696
- Kırkgöz, Y., & Ağçam, R. (2012). Investigating the written assessment practices of Turkish teachers of English at primary education. *The Journal of Language and Linguistic Studies*, 8(2), 119-136.
- Kızıllı, V. (2019). Peer assessment in EFL writing classes. *The Journal of International Lingual Social and Educational Sciences*, 5(1), 101-111. doi: 10.34137/jilses.560099
- Klenowsky, V. (2002). *Developing portfolios for learning and assessment. Processes and principles*. London/New York: Routledge Falmer.
- Knight, P. (2001). *Formative and summative, criterion and norm-referenced assessment*. York. LTSN Generic Centre, Assessment Series No. 7.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275–304. doi:10.1177/0265532208101008
- Koh, K. H. (2017). *Authentic assessment*. In *Oxford research encyclopedia of education*. Oxford: Oxford University Press.
- Korkmaz, I. (2006). Yeni ilköğretim programının öğretmenler tarafından değerlendirilmesi. [Evaluation of new primary education program by teachers] *Ulusal Sınıf Öğretmenliği Kongresi*, 2,249-260. Ankara: Kök Yayıncılık.
- Kwako, J. (2003). A brief summary of traditional and alternative assessment in the college classroom /online/. Retrieved on 20th February 2014 from [http:// www.stat.wisc.edu/~nordheim/Kwako_assessment4.doc](http://www.stat.wisc.edu/~nordheim/Kwako_assessment4.doc).

- Lam, R., & Lee, I. (2010). Balancing the dual functions of portfolio assessment. *ELT Journal*, 64(1), 54-64. doi:10.1093/elt/ccp024
- Lambert, D. & Lines, D. (2013). *Understanding assessment: Purposes, perceptions, practice*. New York, NY: Routledge.
- Larsen-Freeman (1986). *Techniques and principles in language teaching*. Oxford: Oxford University Press.
- Larsen-Freeman, D. (2000). *Techniques and principles in language teaching* (2nd Ed.). New York: Oxford University Press.
- Larsen-Freeman, D. (2003). *Techniques and principles in language teaching*. (2nd Ed.). China.
- Lee, I., & Mak, P. (2014). Assessment as learning in the Language Classroom. *Assessment and Learning* 3, 66-78.
- Letina, A. (2015). Application of traditional and alternative assessment in science and social studies teaching. *Croatian Journal of Education*, 17(1)137-152. doi: 10.15516/cje.v17i0.1496
- Leung, C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly*, 40 (1), 211-234. doi:10.2307/40264517
- Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (8th Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Littlewood, W. (2007). Communicative and task-based language teaching in East Asian classrooms. *Language Teaching*, 40(3), 241-249. doi:10.1017/S0261444807004363
- Lizasoain, A. & Zárate, A. O. (2014). Alternative evaluation of a traditional oral skill assessment tool in an English teaching program. *The LEC Journal*, 1, 21–36.
- Lizzio, A. & Wilson, K. (2013). First year students' appraisal of assessment tasks: Implications for efficacy, engagement, and performance. *Assessment and Evaluation in Higher Education*, 38(4), 389-406 doi: 10.1080/02602938.2011.637156
- Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education*, 20(5), 523-535. doi: 10.1016/j.tate.2004.04.008
- Maclellan, E. (2004). How convincing is alternative assessment for use in higher education? *Assessment & Evaluation in Higher Education*, 29(3), 311–321. doi:10.1080/0260293042000188267
- Marshall, H. H. (1992). *Redefining student learning: Roots of educational change* Norwood, NJ, Ablex.

- McDonald, B., & Boud, D. (2003). The Impact of Self-assessment on Achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, 10(2), 209–220. doi: <https://doi.org/10.1080/0969594032000121289>.
- McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction* (2 Ed.). MA: Pearson Education.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40–49.
- McNamara, T. (2001). Editorial: Rethinking alternative assessment. *Language Testing*, 18 (4), 329–332.
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, 168 (1), 1-5. doi: 10.1501/Dilder_0000000237
- Medfouni, I. (2014). *The impact of self and peer assessment on EFL students' essay writing skills. The Case of third-year L.M.D students of English*. Unpublished master thesis. University of Oum El Bouaghi, Algeria.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11. doi: <https://doi.org/10.2307/1175249>.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded Sourcebook*. (2nd Ed.). Thousand Oaks, CA: Sage.
- Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15, 118-141. doi: 10.1016/j.jslw.2006.01.003
- Ming, T. Y. K. (2002). Assessment strategies that maximize our EPCL students' potential. *Student Assessment and Feedback*, 3, 11-12.
- Ministry of National Education. (2001). *2001 yılı başında milli eğitim. (Ministry of National Education in 2001)*. Retrieved from <http://www.meb.gov.tr/stats/ist2001/Bolum6s1.htm> on January 8, 2010.
- Mirici, İ. H., & Kavaklı, N. (2017). Teaching the CEFR oriented practices effectively in the M.A. program of an ELT department in Turkey. *International Online Journal of Education and Teaching (IOJET)*, 4(1), 74-85. Retrieved from <http://iojet.org/index.php/IOJET/article/view/159/151>.
- MoNE. (2001). *National education at the beginning of 2002*. Republic of Türkiye, MoNE Research, Planning and Coordination Board. Retrieved in December, 2005, retrieved from <http://www.meb.gov.tr>.
- Moqbel, M. S. S. (2015). *Alternative assessment in EFL classroom: How effective?* Egypt: The Nile TESOL/AUC Conference.
- Moon, J. (1999). *Reflection in learning and professional practice*. Kogan Page, London.

- Moon, J.(2004). *A handbook of reflective and experiential learning: Theory and practice*. London: Routledge Falmer.
- Moon, J. (2006). *Learning journals, a handbook for reflective practice and professional development*. 2nd ed. London: Routledge.
- Moradan, A., & Hedayati, S.N. (2012). The impact of portfolios and conferencing on Iranian EFL learners' writing skill. *Journal of English Language Teaching and Learning*, 3(8), 115-141.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research, and Evaluation*, 7(10), 1-6. doi: <https://doi.org/10.7275/q7rm-gg74>.
- Mullin, J. A. (1998). Portfolios: Purposeful collections of student work. *New Directions for Teaching and Learning*, 74, 79-87.
- Nasab, F. G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, 2(6), 165-178.
- Nezakatgoo. B. (2011) Portfolio as a Viable Alternative in Writing Assessment. *Journal of Language Teaching and Research*, 2(4), 747-756. doi: <https://doi.org/10.4304/jltr.2.4.747-756>.
- Norris, J., Brown, J., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments* (Technical Report #18). Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center.
- Obeiah, S. F., & Bataineh, R.F. (2016). The effect of portfolio-based assessment on Jordanian EFL learners' writing performance. *Bellaterra Journal of Teaching & Learning Language & Literature*, 9(1), 32-46. doi: <http://dx.doi.org/10.5565/rev/jtl3.629>
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for language teachers*. Boston: Addison Wesley.
- Ökten, A. (2009). *Effects of formative assessment application on students' language Proficiency in language learning in EFL context: A case study*. Unpublished master thesis, University of Çukurova, Adana.
- Özdemir, S. M. (2010). İlkogretim öğretmenlerinin alternatif olcme ve degerlendirme araclarina iliskin yeterlikleri ve hizmet içi egitim ihtiyaclari. [Competencies of elementary school teachers on alternative assessment tools and in-service training needs]. *Turk Egitim Bilimleri Dergisi*, 8(4), 787-816.
- Özer, O., & Tanriseven, I. (2016). The effect of portfolio-based writing assessment on the development of writing skills of EFL students. *International Online Journal of Educational Sciences*, 8 (3), 35-45. doi: 10.15345/iojes.2016.03.004
- Özusu, N. (2018). *An evaluation of the performance assessment component of an EFL preparatory school program*. Unpublished master thesis, Yeditepe University, Istanbul.

- Paker, T. (2006). Çal bölgesindeki okullarda İngilizce öğretiminin sorunları ve çözüm önerileri. *Çal Sempozyumu*, 684-690. Retrieved from <http://hdl.handle.net/11499/3760>
- Paulson, F.L., Paulson, P.R., & Meyer, C.A. (1991). What makes a portfolio? *Educational Leadership*, 48 (5), 60-63.
- Perlman, C. (2003). *Performance assessment: Designing appropriate performance tasks and scoring rubrics*. U.S. Department of Education.
- Pesen, S. N. (2016). *Using drama and music portfolio as an alternative assessment for developing ELT learners' speaking skills*. Unpublished master thesis. Gazi University, Ankara.
- Piaget, J. (1973). *Psychology and epistemology: Towards a theory of knowledge*. Record: Rio de Janeiro.
- Piri, B. (2022). *A comparative investigation of English preparatory school students' and instructors' perceptions of using e- portfolio as an alternative assessment tool at university*. Unpublished master thesis. Yeditepe University, Istanbul.
- Poehner, M. E. (2005). *Dynamic assessment of advanced L2 learners of French*. Unpublished doctoral thesis. The Pennsylvania State University, University Park, PA.
- Poehner, M. E., & van Compernelle, R. A. (2011). Frames of interaction in dynamic assessment: Developmental diagnoses of second language learning. *Assessment in Education: Principles, Policy & Practice*, 18(2), 183-198. doi:10.1080/0969594X.2011.567116.
- Poehner, M. E., & Inbar-Lourie, O. (2020). An epistemology of action for understanding and change in L2 classroom assessment: The case for praxis. In M. E. Poehner, & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment* (pp. 1-20). Springer.
- Popham, W. J. (2006). Needed: A doze of assessment literacy. *Educational leadership*, 63(6), 84-85.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4-11. doi: 10.1080/00405840802577536
- Prinsloo, P., Slade, & S. Galpin, F. (2011). A phenomenographic analysis of student reflections in online learning diaries. *Open Learning: The Journal of Open, Distance and E Learning*, 26(1), 27-38. doi: <https://doi.org/10.1080/02680513.2011.538562>
- Rea-Dickens, P. (2000). *Classroom assessment*. In T. Hedge (Ed.), *Teaching and learning in the language classroom*, (pp. 375-401). Oxford, UK: Oxford University Press.
- Reeves, T. C. (2000). Alternative assessment approaches for online learning environments in higher education. *Journal of Educational Computing Research*, 23(1), 101-111. doi: <https://doi.org/10.2190/GYMQ-78FA-WMTX-J06C>.

- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472-482. doi: 10.1016/j.tate.2010.09.017
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511667220>.
- Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667190>.
- Richards, J. J., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd Ed). (p.204). New York. Cambridge University Press.
- Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics* . (3rd Ed.) New York: Longman.
- Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th Ed). London: Longman Pearson Education.
- Roberts, T. (2006). *Self-, peer-, and group assessment in e-learning*. United States of America: Information Science Publishing.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, 11(10), 1-13 doi: <https://doi.org/10.7275/9wph-vv65>.
- Rossmann, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review*, 9(5), 627–643. doi: <https://doi.org/10.1177/0193841X8500900505>.
- Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing: The Art of hearing data* (2nd Ed.). Thousand Oaks, CA: Sage.
- Russel, M. K., & Airasian, P. W. (2012). *Classroom assessment: Concepts and applications* (7th Ed.). New York: McGraw-Hill.
- Safari, M., & Koosha, M. (2016). Instructional efficacy of portfolio for assessing Iranian EFL learners' speaking ability. *English Language Teaching*, 9(3), 102—116. doi: 10.5539/elt.v9n3p102
- Salkind, N. J. (2006). *Tests & measurement for people who (think they) hate tests and measurement*. Thousand Oaks: Sage Publications.
- Serafini, F. (2001). Three paradigms of assessment: Measurement, procedure, and inquiry. *The Reading Teacher*, 54(4) 384-393.
- Shakoori, A. M. (2022). *Alternative assessments and English language enhancement: Perspectives of Saudi EFL undergraduate students*. Unpublished doctoral thesis, University of Arizona, Tucson, USA.

- Shepard, L.A. & Dougherty, K.C. (1991). *Effects of high stakes testing on instruction*. Paper presented at the annual meeting of the American Educational Chicago.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14. doi: <https://doi.org/10.3102/0013189X029007>.
- Sheppard, K., & Stoller, F. L. (1995). Guidelines for the integration of student projects in ESP classrooms. *English Teaching Forum*, 33(2), 10-15.
- Shohamy, E. (1982). Affective considerations in language testing. *The Modern Language Journal*, 66(1), 13-17. doi: <https://doi.org/10.2307/327810>
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331-345. doi: [https://doi.org/10.1016/S0191-491X\(98\)00020-0](https://doi.org/10.1016/S0191-491X(98)00020-0).
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Pearson.
- Sidek, H. M. (2012). EFL reading instruction: Communicative task-based approach. *International Journal of Instruction*, 5(2), 109-128. Retrieved from http://www.e-iji.net/dosyalar/iji_2012_2_7.pdf.
- Simkin, M. G., & Kuechler, W. L. (2005). Multiple-choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education*, 3 (1), 73-97. doi:10.1111/j.1540-4609.2005.00053.x
- Simonson, M., Smaldino, S, Albright, M. & Zvacek, S. (2009). *Teaching and learning at a distance: Foundations of Distance Education*, (4th Ed.) Upper Saddle River, NJ: Prentice-Hall.
- Solomon, P.G. (2002). *The assessment bridge - positive ways to link tests to learning, standards and curriculum improvement*. Thousand Oaks, CA: Corwin Press
- Song, B., & August, B. (2002). Using portfolios to assess the writing of ESL students: A powerful alternative? *Journal of Second Language Writing*, 11(1), 49-72. doi:10.1016/S1060-3743(02)00053-X
- Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328. doi:10.1177/003172170508700414
- Stiggins, R.J. (2008). Assessment manifesto: A call for the development of balance assessment systems. *A position paper published by the ETS Assessment Training Institute*, Portland, Oregon.
- Suwaed, H. (2018). EFL Students' perceptions of using portfolio assessments in the writing classroom: The case of Libyan undergraduate second year students. *Journal of Studies in Education*, 8(2), 144–156. doi:10.5296/jse.v8i2.13152
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

- Takkaç Tulgar, A. (2018). Development of a scale measuring attitudes towards English lesson: A study of reliability and validity. *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8 (15), 233-244. doi: 10.29029/busbed.354812
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2). 154-155. doi: 10.1093/eltj/ccj030
- Tedick, D.J., & Klee, C.A. (1998). Alternative assessment in the language classroom. In G. S. Burkart (Ed.), *Modules for the professional preparation of teaching assistants in foreign languages*. Washington, DC: Center for Applied Linguistics.
- Tierney, R. J., M. A. Carter, & L. E. Desai (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher-Gordon Publishers.
- Tompkins, P.K. (2001). The Role playing. Simulation. *TESL*, 4(8).
- Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20-27. doi:10.1080/00405840802577569
- Trochim, W. M. K. (2005). *Research methods: The concise knowledge base*. Atomic Dog Publishing, Inc.
- Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment. *CRILE Working Papers*, 1-23.
- Tsang , W.K. (2003). Journaling from internship to practice teaching. *Reflective Practice*, 4(2): 221 – 40. doi:10.1080/14623940308269
- Tsui, A. B. M., & Tollefson, J. W. (2007). Language policy and the construction of national cultural identity. In A. B. M. Tsui & J. W. Tollefson (Eds.), *Language policy, culture, and identity in Asian contexts* (pp. 1-21). London: Lawrence Erlbaum Associates.
- Underhill, N. (1992). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- Ur, P. (1996). *A course in language teaching: practice and theory*. Cambridge: Cambridge University Press.
- Valencia, S. W., & Calfee, R. (1991). The development and use of literacy portfolios for students, classes, and teachers. *Applied Measurement in Education*, 4(4), 333-345.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* Cambridge, Mass.: Harvard University Press.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1987). Thinking and speech. In R.W. Rieber & A.S. Carton (Eds.), *The collected works of L.S. Vygotsky, Volume 1: Problems of general psychology* (pp. 39–285). New York: Plenum Press. (Original work published 1934.)

- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13 (3), 334-354.
- Weigle, S. (2005). *Assessing writing*. Cambridge: Cambridge University Press.
- Wikström, N. (2008). *Alternative assessment in primary years of international baccalaureate education*. Retrieved from <https://www.diva-portal.org/smash/get/diva2:199424/FULLTEXT01.pdf>
- Williams, P. (2008). Assessing context-based learning: Not only rigorous but also relevant. *Assessment & Evaluation in Higher Education*, 33, 4, 395-408. doi: <https://doi.org/10.1080/02602930701562890>.
- Willis, J. (1981). *Teaching English through English*. New York: Longman.
- Wolf, K., & Siu-Runyan, Y. (1996). Portfolio purposes and possibilities. *Journal of Adolescent & Adult Literacy*, 40, 30-37.
- Woolfolk, A. (2005). *Educational psychology. Active learning edition*. Pearson Allyn& Bacon: Boston.
- Yıldırım, R. & Orsdemir, E. (2013). Performance tasks as alternative assessment for young EFL learners: Does practice match the curriculum proposal. *International Online Journal of Educational Sciences*, 5(3), 562-574.
- Yurdabakan, I., & Erdogan, T. (2009). The effect of portfolio assessment on reading, listening, and writing skills of secondary school prep class students. *The Journal of International Social Research*, 9(2), 526-538.
- Yüce, E., & Mirici, İ. H. (2019). A qualitative inquiry into the application of 9th grade EFL program in terms of the CEFR. *Journal of Language and Linguistic Studies*, 15(3), 1171–1187. doi:<https://doi.org/10.17263/jlls.631560>
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90-102. doi: <https://doi.org/10.1016/j.asw.2018.02.004>.

APPENDICES

Appendix A

Attitude Scale towards Learning English

Dear participant,

The aim of this study, which I conducted as part of my PhD thesis, is to analyze the attitudes of Pamukkale University School of Foreign Languages students towards learning English in general. Your answers will help English language teachers and pedagogues to understand the needs of English language learners and to overcome the problems and difficulties they may encounter in the English language. Your answers to the survey questions will not be shared with third parties and your identity will remain confidential. Participation in the survey is voluntary and you can leave the study at any time. Thank you for your participation in the survey.

PhD Student: Hüsni GÜMÜŞ

Supervisor: Prof. Dr. Turan PAKER

Department:

Preparatory School Education:

Optional/Compulsory

(Write here)

(Underline here)

I agree to participate in the survey/study. _____ (Signature)

	<i>Dear participant, In this scale, there are items to measure your attitudes towards English lesson. strongly agree (5) agree (4) neutral (3) disagree (2), strongly disagree (1) After reading each item carefully, tick the most appropriate option stated above (X).</i>	Completely Agree 5	Agree 4	Neutral 3	Disagree 2	Completely Disagree 1
1	Learning English is a pleasure for me.					
2	Learning English is important for me.					
3	I think that learning English increases my general culture.					
4	Learning English is fun for me.					
5	I think that every student should learn English.					
6	I look forward to English lessons.					

		Completely Agree 5	Agree 4	Neutral 3	Disagree 2	Completely Disagree 1
7	I think that learning English helps me keep up with the globalizing world.					
8	I think that learning English contributes to my awareness of the similarities/differences between languages.					
9	I don't have difficulty in learning English.					
10	I think the "Writing Skills Course" improves my writing skills.					
11	I think the "Speaking Skills Course" improves my speaking skills.					
12	I would take English preparatory education even if it is optional.					
13	I think that learning English improves my perspective towards the mother tongue.					
14	I think that learning English is an opportunity to go abroad.					
15	I think that "Core Language Course" has increased my knowledge of grammar.					
16	I enjoy reviewing topics while learning English.					
17	I think that learning English increases my self-confidence.					
18	I think that learning English increases my knowledge of the features of language use.					
19	I try to watch foreign TV series/movies in English to improve my English proficiency.					
20	I think the "Reading Course" improves my reading skills.					
21	English lesson hours at the preparatory school should be increased.					
22	I think that learning English improves my ability to express myself.					

		Completely Agree 5	Agree 4	Neutral 3	Disagree 2	Completely Disagree 1
23	I try to think in English to improve my English proficiency.					
24	I think the“Listening Skills Course” improves my listening skills.					
25	To improve my English proficiency, I try to write down the expressions I think of in English.					
26	I think that learning English improves my perspective towards foreign languages.					

Takkaç Tulgar, A. (2018). Development of A Scale Measuring Attitudes Towards English Lesson: A Study of Reliability and Validity . *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* , 8 (15) , 233-244 . DOI: 10.29029/busbed.354812.

İngilizce Öğrenmeye Yönelik Tutum Ölçeği

Değerli katılımcı,
Doktora tezim kapsamında yaptığım bu çalışmanın hedefi, Pamukkale Üniversitesi Yabancı Diller Yüksekokulu öğrencilerinin, genel olarak İngilizce öğrenmeye yönelik tutumlarını incelemektir. Cevaplarınız, İngilizce öğretmenlerine ve pedagoglarına, İngilizce öğrenenlerin ihtiyaçlarını anlamaları ve onların İngilizce dilinde karşılaşabilecekleri sorunların ve güçlüklerin üstesinden gelebilmeleri konusunda yardımcı olacaktır. Anket sorularına vereceğiniz cevaplar, üçüncü şahıslarla paylaşılmayacak ve kimliğiniz gizli kalacaktır. Ankete katılım gönüllülük esasına dayalı olup, çalışmadan istediğiniz zaman ayrılabilirsiniz. Ankete katılımınız için teşekkür ederiz.

Öğrenci: Hüsnü GÜMÜŞ

Danışman: Prof. Dr. Turan PAKER

Bölümünüz:

Hazırlık eğitimi: İsteğe bağlı/Zorunlu (Altını

çiziniz)

Ankete/çalışmaya katılmayı kabul ediyorum. _____ (İmzalayınız)

	<i>Değerli katılımcı, Bu ölçekte İngilizce dersine yönelik tutumlarınızı ölçmek üzere maddeler yer almaktadır. Bu maddelerin her birine yönelik; tamamen katılıyorum (5) katılıyorum (4) kararsızım (3) katılmıyorum (2), hiç katılmıyorum (1) şeklinde beş seçenek verilmiştir. Her maddeyi dikkatlice okuduktan sonra size en uygun seçeneği (X) ile işaretletiniz.</i>	Tamamen Katılıyorum 5	Katılıyorum 4	Kararsızım 3	Katılmıyorum 2	Hiç katılmıyorum 1
1	İngilizce öğrenmek benim için bir zevktir.					
2	İngilizce öğrenmek benim için önemlidir.					
3	İngilizce öğrenmenin, genel kültürümü artırdığımı düşünüyorum.					
4	İngilizce öğrenmek benim için eğlencelidir.					
5	Her öğrencinin, İngilizce öğrenmesi gerektiğini düşünüyorum.					
6	İngilizce derslerini dört gözle beklerim.					

		Tamamen Katılıyorum 5	Katılıyorum 4	Kararsızım 3	Katılmıyorum 2	Hiç katılmıyorum 1
7	İngilizce öğrenmenin, küreselleşen dünyaya ayak uydurmama yardımcı olduğunu düşünüyorum.					
8	İngilizce öğrenmenin, diller arasındaki benzerlik/farklılıkların bilincine varmama katkı sağladığını düşünüyorum.					
9	İngilizce öğrenirken zorlanmam.					
10	“Yazma Becerileri” dersinin, yazma becerimi geliştirdiğini düşünüyorum.					
11	“Konuşma Becerileri” dersinin, konuşma becerimi geliştirdiğini düşünüyorum.					
12	İngilizce hazırlık eğitimi, isteğe bağlı olsa da alırım.					
13	İngilizce öğrenmenin, ana dile karşı bakış açımı geliştirdiğini düşünüyorum.					
14	İngilizce öğrenmenin, yurt dışına gidebilmek için bir fırsat olduğunu düşünüyorum.					
15	“Temel İngilizce” (Core language) dersinin, dilbilgisi konusundaki bilgimi artırdığını düşünüyorum.					
16	İngilizce öğrenirken, konuları tekrar etmekten zevk alırım.					
17	İngilizce öğrenmenin, özgüvenimi artırdığını düşünüyorum.					
18	İngilizce öğrenmenin, dillerin kullanım özellikleri konusundaki bilgimi artırdığını düşünüyorum.					
19	İngilizce yeterliğimi geliştirmek için yabancı dizileri/filmleri İngilizce olarak izlemeye çalışırım.					
20	“Okuma Becerileri” dersinin, okuma becerimi geliştirdiğini düşünüyorum.					
21	Hazırlık okulundaki İngilizce ders saatleri artırılmalıdır.					
22	İngilizce öğrenmenin, kendimi ifade edebilme becerimi geliştirdiğini düşünüyorum.					

		Tamamen Katılıyorum 5	Katılıyorum 4	Kararsızım 3	Katılmıyorum 2	Hiç katılmıyorum 1
23	İngilizce yeterliğimi geliştirmek için İngilizce düşünmeye çalışırım.					
24	“Dinleme Becerileri” dersinin, dinleme becerimi geliştirdiğini düşünüyorum.					
25	İngilizce yeterliğimi geliştirmek için aklımdan geçen ifadeleri İngilizce olarak yazmaya çalışırım.					
26	İngilizce öğrenmenin yabancı dile karşı bakış açımı değiştirdiğini düşünüyorum.					

Takkaç Tulgar, A. (2018). Development of A Scale Measuring Attitudes Towards English Lesson: A Study of Reliability and Validity . *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* , 8 (15) , 233-244 . DOI: 10.29029/busbed.354812.

Not: Uygulama Yabancı Diller Yüksekokulunda yapıldığı için bazı maddelerde “İngilizce” sözcüğü yerine “okuma, yazma, dinleme, konuşma dil becerileri ile temel İngilizce” adları ve “İngilizce dersi” yerine “İngilizce öğrenme” sözcüğü kullanılmıştır.

Appendix B

Interview Questions

- 1- What do you think about the evaluation of your performance in the courses with process-based alternative assessment techniques and tools such as portfolio, presentation, debate, video assignment etc.?
- 2- Would you like your performance to be assessed by traditional exams (midterm, final, quiz, etc.) or by the process-based alternative assessment techniques mentioned in question 1? Explain your reasons.
- 3- To what extent were the process-based alternative techniques and assessment tools mentioned in question 1 effective in improving your listening, speaking, writing, and reading skills? Explain for each skill.
- 4- What do you think about the use of checklists and rubrics in the evaluation of alternative assessment tasks mentioned in question item 1?
- 5- What do you think about self-assessment, peer assessment and teacher assessment?
- 6- How do process-based alternative assessment techniques and tools affect your attitude and motivation towards learning English?
- 7- How do traditional assessment tools (midterm, final, quiz) affect your attitude and motivation towards learning English?
- 8- Do you think traditional assessment tools or process-based alternative assessment tools provide more permanent learning? Explain the reasons in detail.

Görüşme Soruları

- 1- Derslerdeki performansınızın, portfolio (dosya) , presentation (sunum) debate (tartışma), video ödevi vs. gibi sürece dayalı alternatif ölçme teknikleri ve araçları ile değerlendirilmesi hakkında ne düşünüyorsunuz?
- 2- Performansınızın tek seferde yapılan geleneksel sınavlar (vize, final, quiz vs.) ile mi, yoksa sürece dayalı şekilde 1. maddede belirtilen tekniklerle mi değerlendirilmesini istersiniz? Sebeplerini açıklayınız.
- 3- Dinleme, konuşma, yazma ve okuma becerilerinizi geliştirmede, 1. Maddede belirtilen sürece dayalı alternatif teknikler ve ölçme araçları ne derecede etkili oldu? Her beceri için açıklayınız.
- 4- 1. Maddede belirtilen alternatif ölçme teknikleri ve aktivitelerin değerlendirilmesinde rubric kullanılması hakkında ne düşünüyorsunuz?
- 5- Öz değerlendirme, akran değerlendirmesi ve öğretmen değerlendirmesi uygulamaları hakkında ne düşünüyorsunuz?
- 6- İngilizce öğrenmeye yönelik tutumunuzu ve motivasyonunuzu sürece dayalı alternatif ölçme araçları nasıl etkiliyor?
- 7- İngilizce öğrenmeye yönelik tutumunuzu ve motivasyonunuzu tek seferde yapılan geleneksel ölçme araçları (vize, final, quiz) nasıl etkiliyor?
- 8- Sizce geleneksel ölçme araçları mı yoksa sürece dayalı alternatif ölçme araçları mı daha kalıcı öğrenme sağlıyor? Sebeplerini detaylı açıklayınız.

Appendix C

Yansıtıcı Günlük Soruları (Konuşma Becerileri Dersi)

Reflective Journal Questions (Speaking Skills Course)

Öğrenci Adı/Soyadı(*Student Name/Surname*):

Sınıf (Class):

Tarih (Date):

- 1- Bu haftaki Konuşma Becerileri dersinde ne tür aktiviteler yaptınız? Bu aktivitelerin hoşunuza giden ve gitmeyen yönleri nelerdi?

What kind of activities did you do in this week's Speaking Skills course? What did you like and dislike about these activities?

- 2- Bu haftaki Konuşma Becerileri dersinin hedefleri “**Nesli tükenme tehlikesi yaşayan çeşitli türler hakkında bilgi edinmek ve bu sorun üzerine tartışarak çözümler üretmek**” şeklindeydi. Bu hedeflerden hangilerine ne derecede ulaştınız?

The objectives of this week's Speaking Skills course were “To learn about various species in danger of extinction and to produce solutions by discussing this problem.” To what extent did you achieve these objectives?

- 3- Bu haftaki Konuşma Becerileri dersi daha önce aldığınız konuşma derslerinden farklı mıydı? Sebeplerini detaylı şekilde belirtiniz.

Was this week's Speaking Skills course different from the speaking lessons you have taken before? State the reasons in detail.

- 4- Bu haftaki Konuşma Becerileri dersinin İngilizce öğrenmeye karşı genel tutumunuza etkisi oldu mu? Detaylı belirtiniz.

Did this week's Speaking Skills course have any effect on your general attitude towards learning English? Please specify in detail.

Yansıtıcı Günlük Soruları (Yazma Becerileri Dersi)

Reflective Journal Questions (Writing Skills Course)

Öğrenci Adı/Soyadı(*Student Name/Surname*):

Sınıf (Class):

Tarih (Date):

- 1- Bu haftaki Yazma Becerileri dersinde ne tür aktiviteler yaptınız? Bu aktivitelerin hoşunuza giden ve gitmeyen yönleri nelerdi?

What kind of activities did you do in this week's Writing Skills course? What did you like and dislike about these activities?

- 2- Bu haftaki Yazma Becerileri dersinin amacı “**İyi düzenlenmiş bir fikir paragrafı yazmak**” şeklindeydi. Bu hedefe ne derecede ulaştınız?

*The objective of this week's Writing Skills course was “**To write a well-organized opinion paragraph .**” To what extent did you achieve this objective?*

- 3- Bu haftaki Yazma Becerileri dersi daha önce aldığınız yazma derslerinden farklı mıydı? Sebeplerini detaylı şekilde belirtiniz.

Was this week's Writing Skills course different from the writing lessons you have taken before? State your reasons in detail.

- 4- Bu haftaki Yazma Becerileri dersinin İngilizce öğrenmeye karşı genel tutumunuza etkisi oldu mu? Detaylı belirtiniz.

Did this week's Writing Skills course have any effect on your general attitude towards learning English? Please specify in detail.

Yansıtıcı Günlük Soruları (Okuma Becerileri Dersi)

Reflective Journal Questions (Reading Skills Course)

Öğrenci Adı/Soyadı(*Student Name/Surname*):

Sınıf (Class):

Tarih (Date):

1- Bu haftaki Okuma Becerileri dersinde ne tür aktiviteler yaptınız? Bu aktivitelerin hoşunuza giden ve gitmeyen yönleri nelerdi?

What kind of activities did you do in this week's Reading Skills course? What did you like and dislike about these activities?

2- Bu haftaki Okuma Becerileri dersinin hedefleri **“Trafik Sıkışıklığı konusu hakkında bir makale okumak ve bu sorunun altında yatan sebepleri ve sorunun çözümü için sunulan yöntemleri okuma alt becerilerini kullanarak belirleyebilmek** şeklindeydi. Bu hedeflerden hangilerine ne derecede ulaştınız?

The objectives of this week's Reading Skills course were “To read an article on the subject of Traffic Congestion and to be able to identify the underlying causes of this problem and the methods offered to solve this problem by using reading sub-skills. To what extent did you achieve these objectives?

3- Bu haftaki Okuma Becerileri dersi daha önce aldığınız okuma derslerinden farklı mıydı? Sebeplerini detaylı şekilde belirtiniz.

Was this week's Reading Skills course different from the reading lessons you have taken before? State the reasons in detail.

4- Bu haftaki Okuma Becerileri dersinin İngilizce öğrenmeye karşı genel tutumunuza etkisi oldu mu? Detaylı belirtiniz.

Did this week's Reading Skills course have any effect on your general attitude towards learning English? Please specify in detail.

Yansıtıcı Günlük Soruları (Dinleme Becerileri Dersi)***Reflective Journal Questions (Listening Skills Course)***

Öğrenci Adı/Soyadı(*Student Name/Surname*):

Sınıf (Class):

Tarih (Date):

- 1- Bu haftaki Dinleme Becerileri dersinde ne tür aktiviteler yaptınız? Bu aktivitelerin hoşunuza giden ve gitmeyen yönleri nelerdi?

What kind of activities did you do in this week's Listening Skills course? What did you like and dislike about these activities?

- 2- Bu haftaki Dinleme Becerileri dersinin hedefleri **“Pandemi üzerine bir seminer dinlemek ve pandeminin özellikleri ve insan hayatı üzerindeki etkilerini analiz edebilmek”** şeklindeydi. Bu hedeflerden hangilerine ne derecede ulaştınız?

The objectives of this week's Listening Skills course were “To listen to a seminar on pandemic and analyze the characteristics of a pandemic and its impact on human life.” To what extent did you achieve these objectives?

- 3- Bu haftaki Dinleme Becerileri dersi daha önce aldığımız dinleme derslerinden farklı mıydı? Sebeplerini detaylı şekilde belirtiniz.

Was this week's Listening Skills course different from the listening lessons you have taken before? State the reasons in detail.

- 4-Bu haftaki Dinleme Becerileri dersinin İngilizce öğrenmeye karşı genel tutumunuza etkisi oldu mu? Detaylı belirtiniz.

Did this week's Listening Skills course have any effect on your general attitude towards learning English? Please specify in detail.

Appendix D

Writing Rubric for B1 and Above Levels

CONTENT (40 Marks)

(33-40) VERY GOOD	Ideas expresses fully , covering all content elements with appropriate elaboration and minimal digression. Completely relevant to the assigned task. Interesting and informative
(26-32) GOOD	Ideas expressed covering all content elements with some elaboration. There may be some minor repetition or digression. Relevant to the task and require minimal effort to read. Informative and somewhat interesting.
(15-25) ADEQUATE	A simple account with little elaboration or with some repetition and digression from the task. One or two content elements may have been ignored. Content may have been covered, however, not very interesting , but monotonous.
(6-14) INADEQUATE	Not enough information. Student is jumping from one point to the other . Noticeable digression and irrelevance to the task. Requires considerable effort to follow.
(3-5) POOR	Totally irrelevant to the assigned task or information is too little to assess.

ORGANIZATION (20 Marks)

(16-20) VERY GOOD	Ideas clearly stated, supported by various examples, facts, or details. Well-organized and developed. Fully cohesive.
(11-15) GOOD	Main ideas stand out but loosely organized or somewhat supported by various examples, facts or details. Still cohesive
(6-10) ADEQUATE	Only topic sentence and some factual information have been expressed. Limited support. Non-fluent. Lack of cohesion.
(3-5) INADEQUATE	Ideas confused or disconnected . No cohesion at all.
(0-2) POOR	Ideas do not communicate . No organization or not enough to assess.

VOCABULARY (20 Marks)

(16-20) VERY GOOD	Effective word choice and appropriate usage fully relevant to the task. A wide range of vocabulary has been used and even there may be idiomatic expressions.
(11-15) GOOD	Quite precise use of vocabulary but still occasional inappropriate usage without obscuring meaning.
(6-10) ADEQUATE	Adequate usage of vocabulary with some hesitation . Some repetitions and searching for a word. Students may not remember some words but replaces with the ones from L1.
(3-5) INADEQUATE	Vocabulary focused on basic objects, places, and common words. Frequent inappropriate usage of words.
(0-2) POOR	Not enough usage of vocabulary to assess.

ACCURACY (15 Marks)

(13-15) VERY GOOD	Good control and confident use of language including complex statements and range of structures. There may be few errors of agreement, tense, number, articles or prepositions.
(9-12) GOOD	Effective but simple constructions including minor problems in complex structures, a few errors of agreement, tense, number, articles or prepositions.
(5-8) ADEQUATE	Major problems in structure and sometimes require careful reading. Meaning is sometimes obscured . Several errors of agreement, tense, number, articles or prepositions.
(3-4) INADEQUATE	Difficult to follow due to frequent grammatical errors. Poor sentence construction or so much translation of syntax from L1
(0-2) POOR	No mastery of sentence structure or not enough information to assess.

MECHANICS (5 Marks)

(4-5) EXCELLENT TO VERY GOOD	Demonstrates mastery of conventions; few errors of spelling, punctuation, capitalization, paragraphing but meaning not obscured .
(2-3) FAIR TO POOR	Frequent errors of spelling, punctuation, capitalization, paragraphing; poor handwriting ; meaning confused or obscured.
(0-1) VERY POOR	No mastery of conventions; dominated by errors of spelling, punctuation, capitalization, paragraphing; handwriting illegible ; or not enough to evaluate.

Hughes, A. (2003). *Testing for language teachers*, Cambridge University Press.

Appendix E (Oral Presentation Rubric)

Non-verbal Skills (5)	Holds attention of entire audience with the use of direct eye contact, seldom looking at notes; Movements seem fluid and help the audience visualize; Student displays relaxed, self-confident nature with no mistakes. <u>5</u>	Consistent use of direct eye contact with audience, but still returns to notes; Made movement or gestures that enhances articulation; Makes minor mistakes but recovers gracefully; displays little or no tension. <u>4</u>	Displayed minimal eye contact with audience, read from notes; very little movement or gestures; displays mild tension; has trouble recovering from mistakes <u>3</u>	No eye contact with audience, read from notes or paper; no animation in gestures, stood rigid; tension and nervousness is obvious; has trouble recovering from mistakes <u>1-2</u>
Speaking Skills (40)	Demonstrates a strong, positive feeling about topic; student uses a clear voice and correct, precise pronunciation of terms, everyone can hear. <u>35-40</u>	Occasionally shows position feelings about topic; Voice is clear, most words are pronounced correctly. Most audience members can hear the presentation. <u>25-34</u>	Shows only little interest in topic being presented; Voice is low, several terms are mispronounced, Audience has trouble hearing presentation. <u>15-24</u>	Shows no interest in topic; student mumbles, incorrectly pronounces terms or speaks too quietly for the majority of audience to hear. <u>1-14</u>
Subject Knowledge (20)	Student demonstrates full knowledge of topic. <u>15-20</u>	Student is at ease with content of presentation <u>10-14</u>	Student seems uncomfortable with Information <u>5-9</u>	Student does not have grasp of information. <u>1-4</u>
Organization (10)	Student presents information in logical, interesting sequence which audience can follow. <u>9-10</u>	Student presents information in logical sequence which audience can follow. <u>6-8</u>	Audience has difficulty following presentation because student jumps around. <u>4-5</u>	Audience cannot understand presentation because there is no sequence of information. <u>1-3</u>
Mechanics (5)	Presentation has no misspellings or grammatical errors. Slides are clear, neat and not too dense. <u>5</u>	Presentation has no more than two misspellings and/or grammatical errors. Slides are plain but readable. <u>3-4</u>	Presentation has several misspellings or grammatical errors. Slides are too dense, or color choices are ineffective. <u>2</u>	Student's presentation has many spelling errors or grammatical mistakes. Slide design is very much in need of improvement. <u>1</u>
Handling of Questions (15)	Easily explains and elaborates on expected questions. <u>13-15</u>	Demonstrates sufficient knowledge of the material to answer expected questions. <u>10-12</u>	Has some difficulty understanding questions or answering beyond a rudimentary level. <u>6-9</u>	Student demonstrates an inability to answer most questions. <u>1-5</u>
Length of Presentation (5)	Presentation was well planned and finished within 1-2 minutes of expected time. <u>5</u>	Presentation was a bit too long or too short, but the student adjusted. <u>4</u>	Presentation was 5-6 minutes too short, or the student had to be encouraged multiple times to stop. <u>3</u>	Presentation was way too short (6 mins) or student had to be stopped for going too long, despite warnings <u>1-2</u>

Adapted from: <https://john.cs.olemiss.edu/hccresearchMethodsformsOralPresentationRubric>

Appendix F
Classroom Debate Rubric

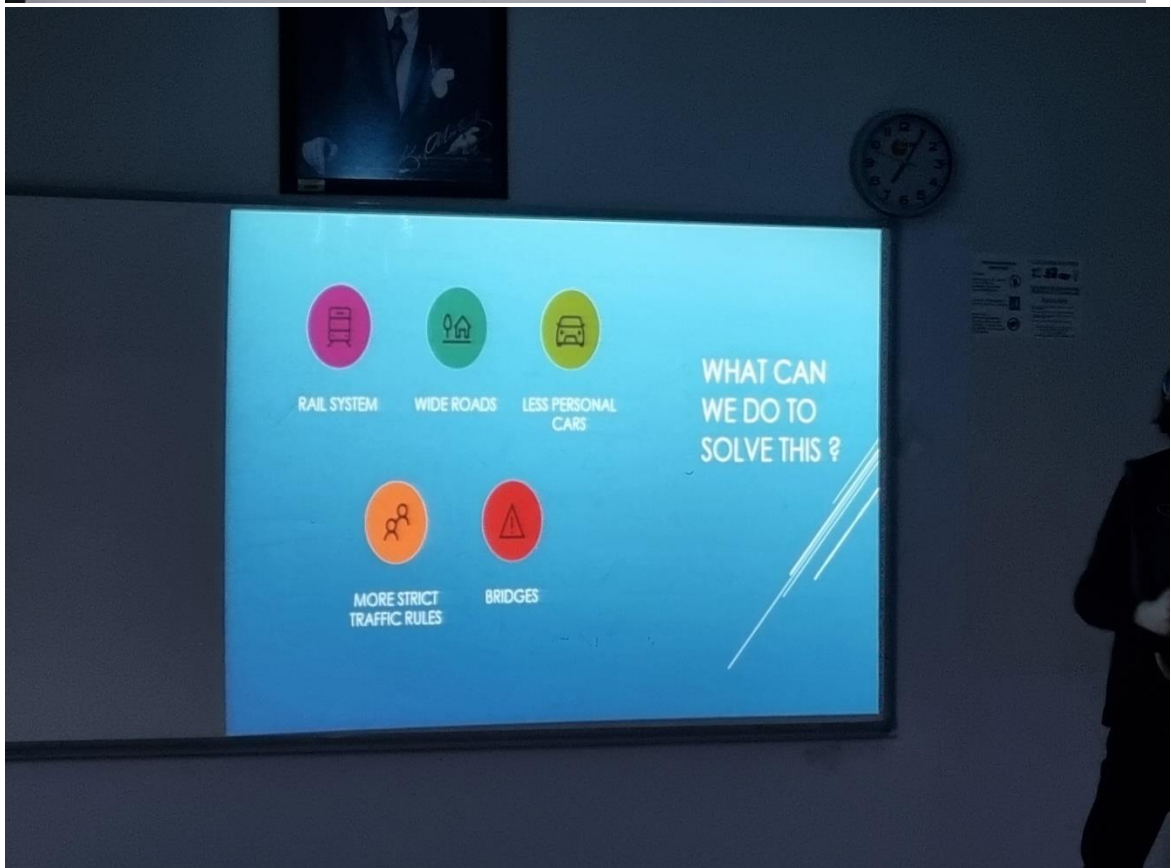
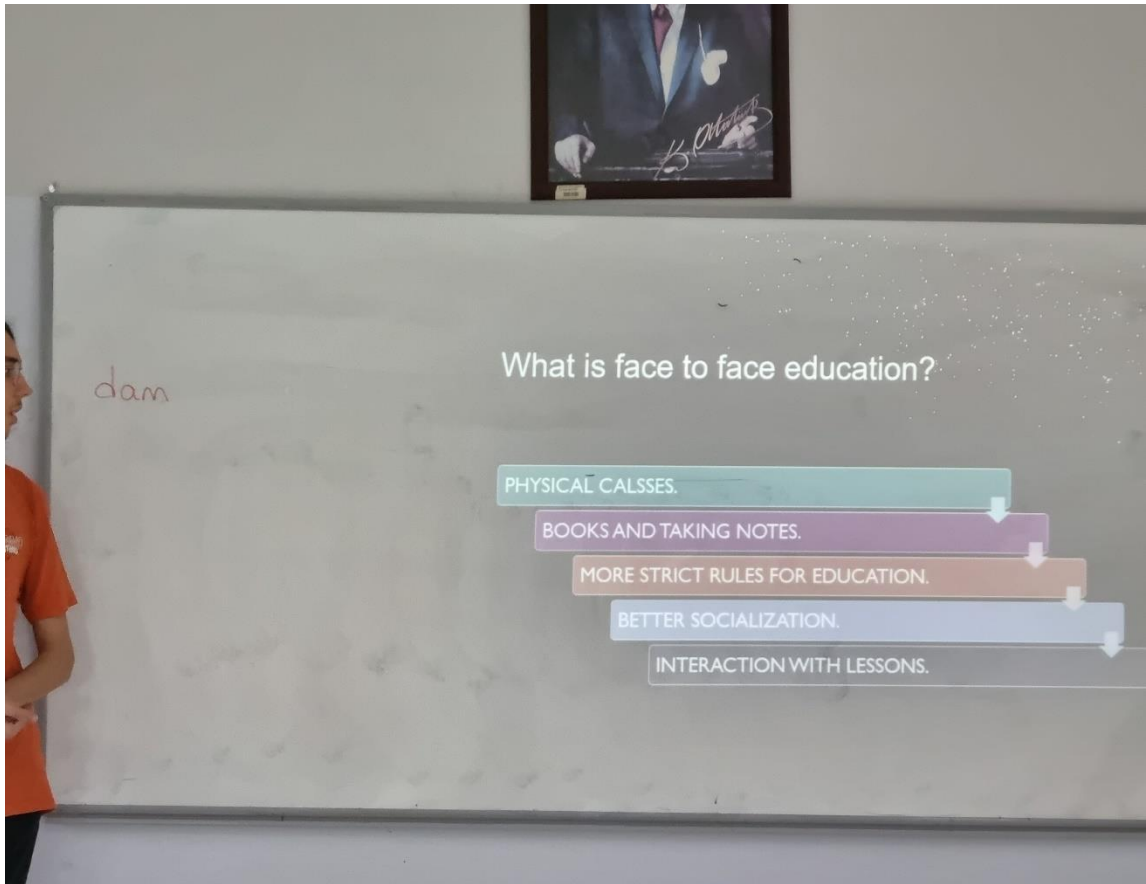
Criteria						Total Points
Respect for Other Team 5	All statements, body language, and responses were respectful and were inappropriate language <u>5</u>	Statements and responses were respectful and used appropriate language, but once or twice body language was not <u>4</u>	Most statements and responses were respectful and in appropriate language, but there was one sarcastic remark <u>3</u>	Statements, responses and/or body language were borderline appropriate. Some sarcastic remarks <u>2</u>	Statements, responses and/or body language were consistently not Respectful <u>1</u>	
Information 20	All information presented in this debate was clear, accurate and Thorough <u>16-20</u>	Most information presented in this debate was clear, accurate and Thorough <u>12-16</u>	Most information presented in the debate was clear and accurate, but was not usually Thorough <u>7-11</u>	Some information was accurate, but there were some minor inaccuracies <u>6-10</u>	Information had some major inaccuracies OR was usually not clear <u>1-5</u>	
Rebuttal 15	All counter-arguments were accurate, relevant and strong <u>13-15</u>	Most counter-arguments were accurate, relevant, and strong <u>10-12</u>	Most counter-arguments were accurate and relevant, but several were weak <u>7-9</u>	Some counter arguments were weak and irrelevant <u>4-6</u>	Counter-arguments were not accurate and/or relevant <u>1-3</u>	
Use of Facts/Statistics 15	Every major point was well supported with several relevant facts, statistics and/or Examples <u>13-15</u>	Every major point was adequately supported with relevant facts, statistics and/or Examples <u>10-12</u>	Every major point was supported with facts, statistics and/or examples, but the relevance of some was Questionable <u>7-9</u>	Some points were supported well, others were not <u>4-6</u>	All points were not Supported <u>1-3</u>	
Organization 15	All arguments were clearly tied to an idea (premise) and organized in a tight, logical fashion <u>13-15</u>	Most arguments were clearly tied to an idea (premise) and organized in a tight, logical fashion <u>10-12</u>	Most arguments were clearly tied to an idea (premise) and organized in a tight, logical fashion <u>7-9</u>	Most arguments were clearly tied to an idea (premise) and organized in a tight, logical fashion <u>4-6</u>	Most arguments were clearly tied to an idea (premise) and organized in a tight, logical fashion <u>1-3</u>	
Understanding of Topic 30	The team clearly understood the topic in depth and presented their information forcefully and Convincingly <u>25-30</u>	The team clearly understood the topic in depth and presented their information with Ease <u>19-24</u>	The team seemed to understand the main points of the topic and presented those with Ease <u>13-18</u>	The team seemed to understand the main points of the topic, but didn't present with ease <u>7-12</u>	The team did not show an adequate understanding of the topic <u>1-6</u>	
Total Points:						

Adapted from: <http://course1.winona.edu/shatfield/air/classdebate.pdf>

Appendix G

Samples of Alternative Assessment Tasks Presented by the Students



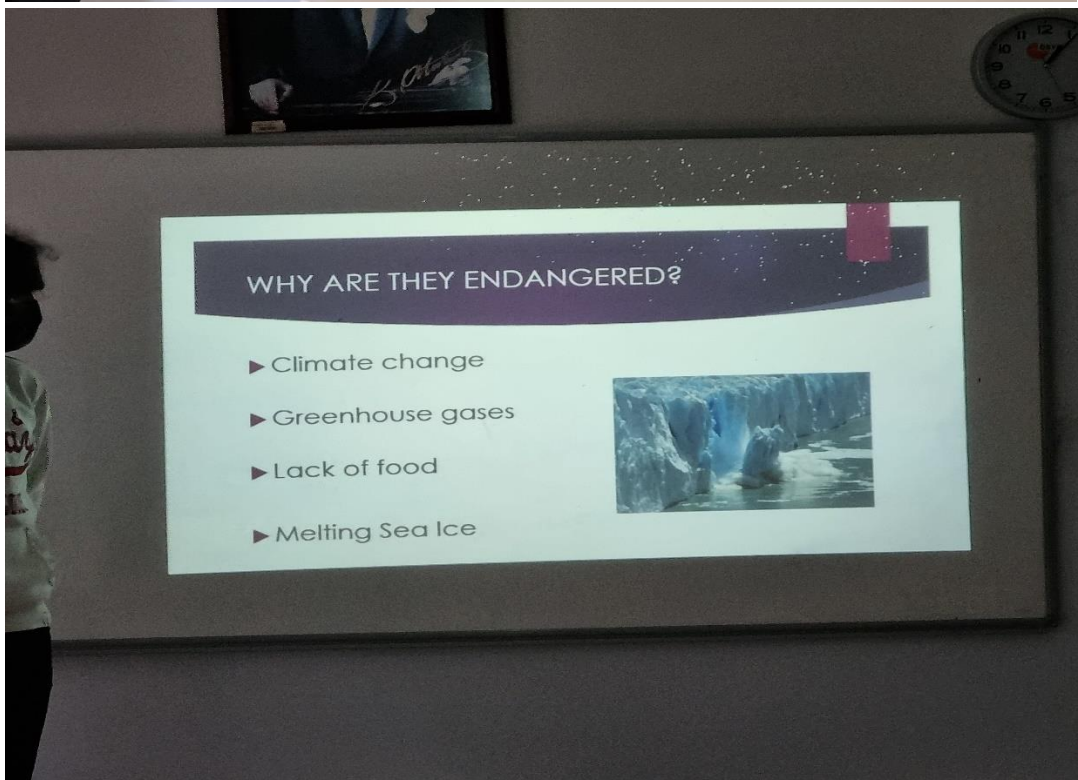




The slide features an illustration of a person at a computer workstation with various educational icons like a globe, books, and a laptop. Below the illustration, the text 'What are the advantages?' is displayed. To the right, a bulleted list outlines the benefits of a flexible learning system.

What are the advantages?

- Flexible schedule allows students to study at anytime
- Students can retake the lesson with recorded lessons
- Homeworks can be done faster and more proficient.
- Teachers can control the situation more efficiently.



Samples of Students' Writings (Pages from the Online Magazine)

EDUCATION

Online education is Great

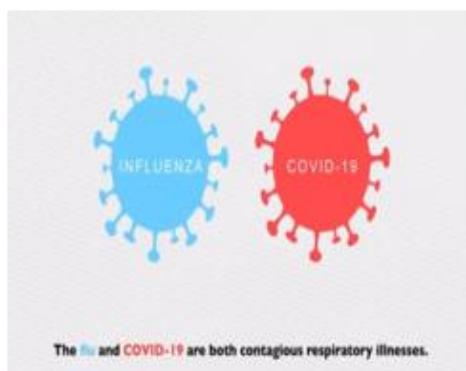
There are three important advantages of online education. Most importantly, online education is easy and comfortable. For example, students don't have to leave their houses to get to school. Also, they don't need to get dressed for school. Secondly, distance doesn't matter in online education. In other words, you can join the lesson no matter where you are. To give an example from my own life, I was in my village during the pandemic. Thanks to online education, I was able to join the lessons easily. In addition, it was possible to submit my assignments through email. Lastly, it is economical to study online. Students don't need to pay for transportation or food. Moreover, they can access to unlimited course materials for free. In short, online education is a great option for students due to these advantages. As a student, I believe that all kinds of education will be delivered online in the future.





Flu and Covid-19

There are many differences between Covid-19 and the flu. The first difference is Covid-19 is a pandemic disease while the flu is an ordinary disease. In other words, Covid-19 has affected the whole world and caused many deaths. However, Flu is observed every year and it is usually not fatal. Secondly, while we need to have more than one vaccine against Covid-19, a single vaccine is sufficient for the flu. For example, I have been vaccinated three times against Covid-19 this year, but I used to get vaccinated only once against the flu. The last difference is while there is a loss of taste and smell due to Covid-19, such a thing does not happen in the flu. According to a recent study, 70% of people who have suffered from Covid-19 still cannot taste or smell as well as they did in the past. In short, Covid-19 and Flu have a lot of differences in several ways. I think, everyone should be vaccinated to protect themselves against these diseases.



Appendix H

Etik Kurulu ve Ölçek Uygulama İzni Belgeleri

(Ethics Committee Approval)

T.C.
PAMUKKALE ÜNİVERSİTESİ
SOSYAL VE BEŞERİ BİLİMLERİ BİLİMSEL ARAŞTIRMA VE YAYIN ETİĞİ KURULU
SAYI: 68282350/22021/G020

Toplantı Tarihi : 3.11.2021
Toplantı Sayısı : 20
Toplantı Saati : 15:00

10.160.1.116
97511
8.11.2021

KARAR 5- Üniversitemiz Eğitim Bilimler Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Doktora Programı 192157001 numaralı öğrencisi Hüsnü GÜMÜŞ'ün "*The Use of Alternative Assessment and Its Effeckt on Sutudents Akademik Achievement and Their Attitudes towards Learning English in Preparatory Schools*" konulu tez çalışmasına yönelik başvuru formunun usul ve etik açıdan verdiği beyan ve ekler tetkik edilmiş olup; proje sahibinin, başvurusunda yer alan bilgi, belge ve taahhütnamelere uygun bilimsel davranışlar sergileyeceği kanaati oluşmuştur. İş bu karar oy birliği ile alınmıştır.

Prof. Dr. Ertuğrul İŞLER
Başkan



T.C.
PAMUKKALE ÜNİVERSİTESİ
Öğrenci İşleri Daire Başkanlığı

Sayı : E-93282220-300-160124
Konu : Ölçek Uygulama İzni (Hüsnü GÜMÜŞ)

25.01.2022

DAĞITIM YERLERİNE

İlgi : 20.01.2021 tarihli ve E-30575850-302.08.01-158959 sayılı yazınız.

Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Doktora Programı 192157001 numaralı öğrencisi Hüsnü GÜMÜŞ'ün, "Hazırlık Okullarında Alternatif Ölçme Uygulamalarının Öğrenci Başarısına ve İngilizce Öğrenme Tutumuna Etkisi" başlıklı tez çalışması kapsamında, Üniversitemiz Yabancı Diller Yüksekokulundan İlgi'de kayıtlı yazı Ek'inde yer alan ölçek aracılığı ile veri toplama talebi Rektörlüğümüzce uygun görülmüştür.

Bilgilerini ve gereğini rica ederim.

Prof. Dr. İbrahim KISAÇ
Rektör Yardımcısı

Ek: İlgi Yazı ve Ekleri (8 Sayfa)

Dağıtım:
Gereği:
Eğitim Bilimleri Enstitüsü

Bilgi:
Yabancı Diller Yüksekokulu