



TC. YEDİTEPE UNIVERSITY

INSTITUTE OF EDUCATION SCIENCES

DEPARTMENT OF ENGLISH LANGUAGE TEACHING



AN EVALUATION OF THE PERFORMANCE ASSESSMENT COMPONENT  
OF  
AN EFL PREPARATORY SCHOOL PROGRAM

NİHAN ÖZUSLU

JUNE, 2018

AN EVALUATION OF THE PERFORMANCE ASSESSMENT COMPONENT  
OF  
AN EFL PREPARATORY SCHOOL PROGRAM

NIHAN ÖZUSLU

MA THESIS  
IN  
DEPARTMENT OF ENGLISH LANGUAGE TEACHING

Supervisor: Assist. Prof. Dr. Zeynep ÇAMLIBEL ACAR

YEDİTEPE UNIVERSITY

JUNE, 2018



YEDİTEPE UNIVERSITY  
INSTITUTE OF EDUCATIONAL SCIENCES DIRECTORATE

THESIS SUBMISSION and APPROVAL FORM

APPROVAL:

Dr. Öğr. Üy. Zeynep C. Canlıbel-Acar  
(Advisor)

[Signature]  
(Signature)

Dr. Öğr. Üy. Evrin Evyik-Aydın  
(Member)

[Signature]  
(Signature)

Dr. Öğr. Üy. Zeynep Kosoğlu  
(Member)

[Signature]  
(Signature)

SUBMITTED BY : Nihan ÖZÜSLÜ  
DATE OF THESIS DEFENSE : 13.06.2018  
DATE OF THESIS APPROVAL : 13.06.2018

## ABSTRACT

### AN EVALUATION OF THE PERFORMANCE ASSESSMENT COMPONENT OF AN EFL PREPARATORY SCHOOL PROGRAM

Nihan ÖZUSLU

Supervisor: Assist. Prof. Dr. Zeynep ÇAMLIBEL ACAR

The purpose of the present study is to evaluate the use of Performance-based Assessment Tasks (PTs) as the supplementary assessment component of the English language program carried out in a preparatory school of a state university from the perspective of the students, instructors and administrators. It explores their expectations before the implementation and their opinions after the implementation on the planning, application, scoring and learning outcomes phases and also its consistency with the current English language program in 2017-2018 academic year. The formative evaluation is based on the Decision/Accountability-Oriented evaluation approach introduced by Stufflebeam (2001). To conduct this mixed-methods study, an open-ended questionnaire is designed and applied to 126 students, 60 instructors and two administrators to collect qualitative data. For the quantitative data, a Likert-scale is designed and applied to the same participants. In addition, opinions and suggestions of the stakeholders for possible future PT practices are gathered via three open-ended questions. The qualitative data is analyzed by means of content and thematic analysis while the data gathered through the scales are analyzed by means of a statistical program to yield means and percentages. Results for the qualitative data collected before the PT implementation indicate that all stakeholders agree on the linguistic

benefits of PTs and its contribution to learning. Students' worries focus on time, effort and difficulty of PT practices whereas instructors embody mainly assessment-related concerns. Administrators expect the PT practices to affect learner autonomy and motivation positively while they are aware of the workload for the instructors. The results of the scale conducted after the implementation reveal that when the tasks are not favored, it is mostly due to its number, content and collecting procedures. The planning, application, scoring, learning outcomes and program consistency phases are evaluated in a moderately positive way by the students while evaluated more positively by the instructors and the administrators. Even though most stakeholders urge some revisions on PTs, a large number of instructors, a great majority of students and the administrators agree on its continuity in the following years. The suggestions made by the stakeholders are presented and taken into consideration to enhance this assessment component in accordance with the current language program. The overall evaluation is discussed with respect to the decision/accountability-oriented evaluation principles. In conclusion, this study provides valuable findings for the field of education particularly for program evaluators, practitioners and researchers in terms of not only assessment but also evaluation.

*Key Words: alternative assessment, performance-based assessment tasks, program evaluation*

## ÖZET

### BİR İNGİLİZCE HAZIRLIK PROGRAMI BÜNYESİNDEKİ PERFORMANSA DAYALI ÖLÇME UYGULAMALARININ DEĞERLENDİRİLMESİ

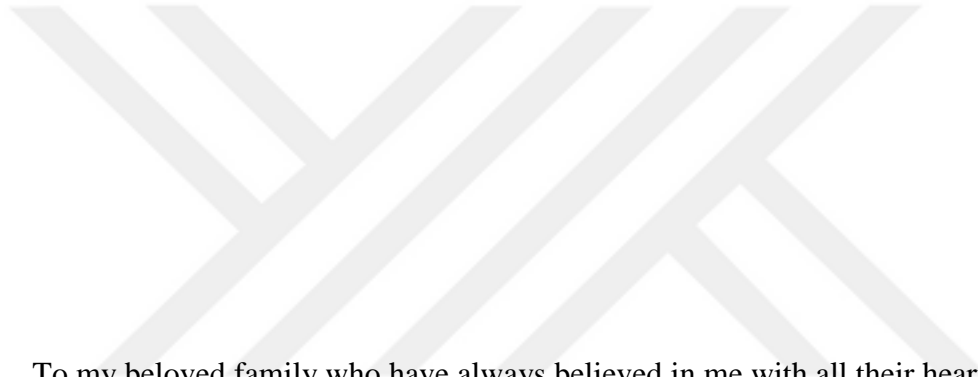
Nihan ÖZUSLU

Tez Danışmanı: Yard. Doç. Dr. Zeynep ÇAMLİBEL ACAR

Bu çalışmanın amacı, Türkiye’deki bir devlet üniversitesi bünyesinde İngilizce yabancı dil eğitimi veren bir hazırlık okulunda öğrencilerin İngilizcedeki başarı ve performansını ölçmek için ağırlıklı olarak kullanılan geleneksel sınav yöntemlerine ek olarak performansa-dayalı ölçme-değerlendirme tekniklerinin uygulanmasını irdelemektir. Bu çalışma ile alternatif ölçme ve değerlendirme yöntemi olarak kabul edilen performans görevlerinin uygulaması öncesinde öğrenci, öğretim görevlisi ve idarecilerin tutumları, uygulama sonrası görüş ve önerileri ve bu alternatif ölçme tekniğinin planlama ve uygulama safhaları ile mevcut dil programına uygunluğu gibi konular değerlendirilmektedir. Dil programının önemli bir parçası olan ölçme-değerlendirmenin bu alternatif tamamlayıcı uygulaması Stufflebeam (2001) tarafından öne sürülen gelişim-odaklı değerlendirme yaklaşımlarından ‘Karar/hesap verebilirlik-odaklı’ değerlendirme prensipleri çerçevesinde irdelenmiştir. Çalışmada veri toplama aracı olarak açık-uçlu sorular ve 5-li Likert ölçeği kullanılmış olup, çalışmaya toplam 129 öğrenci, 60 öğretim görevlisi ve iki idareci katılmıştır. Çalışmanın nicel bulguları içerik ve tematik analiz ile nitel bulguları istatistiksel bir program aracılığıyla analiz edilmiş olup, tüm bu analizler uzman görüşü kontrolünde sunulmuştur. Çalışma sonuçlarına göre uygulama öncesi tüm paydaşlar performans görevlerinin dile ve öğrenime katkısı konusunda hem fikir iken, öğrenciler uygulamanın olası negatif

tarafının fazladan zaman ve çaba harcanması olduğunu belirtmiştir. Öğretim görevlileri ise uygulamanın negatif yönlerinden bahsederken genelde ölçme-değerlendirmeye yönelik endişelerini ifade etmişlerdir. Hazırlık okulu idarecileri ise performans görevlerinden öğrenci otonomisi ve güdülenmesi adına olumlu beklentiler içinde olup bu uygulamanın uygulayıcılar için meşakkatli bir iş olduğunun altını çizmişlerdir. Uygulama sonrası katılımcılara planlama, uygulama, notlandırma, öğrenim kazanımları ve programla uyum aşamalarını değerlendirmeleri için sunulan anketin sonuçlarına göre performans görevlerinin katılımcılar tarafından olumlu karşılandığı ve mevcut ders programıyla örtüştüğü gözlemlenmiştir. Anketin açık-uçlu bölümündeki katılımcı yanıtlarına göre sırasıyla öğrencilerin çoğu, öğretim görevlilerinin büyük çoğunluğu ve idarecilerin her ikisi de performans uygulamalarının özellikle performans görev sayıları, içeriği ve sunma teknikleri gibi bazı değişiklikler yapıldıktan sonra gelecek yıllarda da devam etmesi yönünde fikir beyan etmişlerdir. Çalışma sonunda tüm uygulama süreci Karar/Hesap verebilirlik-odaklı değerlendirme prensipleri çerçevesinde irdelenmiştir. Bu çalışmada sunulan nitel ve nicel bulguların benzer alternatif değerlendirme yöntemlerini uygulamak isteyen kurumlar ve program değerlendiricileri için örnek teşkil etmesi ve dil öğretimine katkı sağlaması ümit edilmektedir.

*Anahtar Kelimeler: alternatif ölçme-değerlendirme, performans görevleri, program değerlendirme*



To my beloved family who have always believed in me with all their heart and soul



## ACKNOWLEDGEMENTS

As my supervisor, as my guide but above all as one of the most precious people I have ever met, I would like to express my gratitude to Assist. Prof. Dr. Zeynep ÇAMLIBEL ACAR for her explicit feedback, support and patience she showed throughout my thesis process.

I would like to thank Assist. Prof. Dr. Zeynep KOÇOĞLU for the guidance she provided any time I had a question on my mind. I owe many thanks to my dear lecturer Assist. Prof. Dr. Evrim EVEYİK AYDIN for her all contributions to me during my master courses and thesis process. I would like to thank Prof. Dr. Hossein FARHADY to share his knowledge in assessment and to encourage me to do what I believe in.

I owe many thanks to my director Prof. Dr. Recep ÇIBİK without whose permission I cannot have my MA experience at all. He not only gave me this opportunity but also encouraged me with his constant personal support. As my colleague, as my helper, as my friend, I cannot thank Mehmet Doğan enough for his contribution, patience and effort he made for me in this tedious period. I am grateful to research assistant Ender UZABACI and Mehmet DOĞAN for sparing time for the statistical analysis of the data.

I would like to thank my colleagues and my friends Nazlı KOÇ, Berrin ÇETİNAVCI, Handehan ÖZGÖRÜR, Sevim AKALAN, Zülkade GAZİ, Zeynep SAKA, Mine ÖZÇİRA, and Sezen BALAMUR for their personal effort as the coordinators of this new application under the investigation of this study. I would like to thank them all to believe in what they are doing and to do their best to make it real. I would also like to thank Billur YILDIRIM, Özge ÖZTÜRK and Emine YILDIRIM

for their help in my data-collection process. I owe many thanks to the whole staff of the School of Foreign Languages where the present study was conducted for their great efforts to administer the procedures of this research.

Not only as a colleague but also as a close friend, I owe many thanks to dear Aliye Evin YÖRÜDÜ for answering my endless questions, for sharing her experiences and first and foremost for encouraging me to keep going under any circumstances. I would like to thank my old friend dear Nevra ÇAKMAK for making her presence felt any time I need support and for being my comrade during this tedious experience as she always did.

My last and deepest thanks go to my dear family; my father and my role model Mehmet Fehmi ÖZUSLU, my devoted mother Ayten ÖZUSLU and my dear brother Korhan ÖZUSLU for their belief in me , sometimes more than me myself, to overcome any obstacle I have come across, for their endless encouragement to do my best in life. Above all I owe them many thanks to be my most powerful advocates in any work I do.

I would like to thank all these precious people for their support and faith to make one of my dreams come true.

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>iii</b>
<b>ÖZET</b> .....	<b>v</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>viii</b>
<b>TABLE OF CONTENTS</b> .....	<b>x</b>
<b>LIST OF TABLES</b> .....	<b>xiv</b>
<b>ABBREVIATIONS</b> .....	<b>xvii</b>
<b>SIGNIFICANT TERMS</b> .....	<b>xviii</b>
<b>CHAPTER 1 INTRODUCTION</b> .....	<b>1</b>
1.1. Background of the Study.....	1
1.2. Statement of the Problem .....	2
1.3. Significance of the Study and Research Questions.....	5
1.4. Overview of the Methodology.....	8
1.5. Limitations of the study .....	8
<b>CHAPTER 2 LITERATURE REVIEW</b> .....	<b>10</b>
2.1. Program Evaluation.....	10
2.1.1. Types and Models of Program Evaluation .....	11
2.1.2. Decision/Accountability-Oriented Approach.....	12
2.1.3. Research on Program Evaluation .....	14
2.2. Assessment .....	17
2.2.1. Assessment, Evaluation and Testing.....	17
2.2.2. Assessment Types .....	18
2.2.3 Language Assessment & Trends in Language Assessment .....	19
2.3. Language Assessment at Higher Education (Preparatory schools) in Turkey.....	21
2.4. Traditional Assessment .....	22
2.4.1. Pros of Traditional Assessment.....	23
2.4.2. Cons of Traditional Assessment.....	25
2.5. Alternative Assessment .....	27
2.5.1. Pros of Alternative Assessment .....	28
2.5.1. Cons of Alternative Assessment .....	30
2.6. The Combination of Traditional and Alternative Assessment .....	32
2.7. Performance-based Assessment.....	34
2.7.1. Performance-based Assessment Tasks.....	37
2.7.2. Pros of Performance-based Assessment .....	40

2.7.3. Cons of Performance-based Assessment.....	42
2.7.4. Research on PBA .....	43
<b>CHAPTER 3 METHODOLOGY .....</b>	<b>47</b>
3.1. Context.....	47
3.2. Participants .....	49
3.2.1. Students .....	50
3.2.2. Instructors.....	50
3.2.3. Administrators .....	51
3.3. General Assessment Procedures in the Context.....	51
3.3.1. Assessment Procedures before the PT implementation (Fall 2016 Semester) .....	52
3.3.2. PT Implementation Trial (Spring 2017 Semester).....	53
3.3.3. Assessment Procedures during the PT implementation (Fall 2017 Semester).....	54
3.4. Current PT implementation in the Context .....	55
3.4.1. The Objective .....	55
3.4.2. Planning .....	56
3.4.3. Performance Assessment Task Descriptions .....	56
3.4.4. Scoring Rubrics.....	59
3.4.6. Assessing PTs.....	68
3.4.7. Targeted Learning Outcomes.....	69
3.4.8. Final Evaluation Meeting with Instructors.....	69
3.5. Data Collection.....	70
3.5.1. Questionnaires with Open-ended Items.....	70
3.5.2. Survey with Likert-type and Open-ended Items.....	71
3.6. Data Collection Procedures.....	74
3.7 Data Analysis.....	75
<b>CHAPTER 4 FINDINGS .....</b>	<b>76</b>
4.1. What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of stakeholders? .....	76
4.1.1. What were the positive and negative aspects of performance assessment tasks before their implementation from the perspective of students? .....	77
4.1.2. What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of instructors? ...	82
4.1.3. What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of administrators? .....	86

4.2. How were the planning, application, scoring, learning outcomes and program consistency phases of performance-based assessment tasks evaluated after their implementation by stakeholders? .....	89
4.2.1. How were the planning, application, scoring, learning outcomes and program consistency phases of PTs evaluated after their implementation by students? .....	89
4.2.2. How were the planning, application, scoring, learning outcomes and program consistency phases of PTs evaluated after their implementation by instructors?.....	95
4.2.3 How were the planning, application, scoring, learning outcomes and program consistency phases of PTs evaluated after their implementation by administrators? .	101
4.3.What were the views on the continuity of the current performance assessment implementation as stated by stakeholders?.....	107
4.3.1 What were the views on the continuity of the current performance assessment implementation as stated by students? .....	108
4.3.2. What were the views on the continuity of the current performance assessment implementation as stated by instructors? .....	114
4.3.3. What were the views on the continuity of the current performance assessment implementation as stated by administrators?.....	120
4.4 Summary of Findings.....	121
<b>CHAPTER 5 DISCUSSION .....</b>	<b>125</b>
5.1. Evaluation of the Evaluation .....	125
5.2. Discussion.....	131
5.2.1. What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of students, instructors and administrators?.....	131
5.2.2. How were the planning, application, scoring, learning outcomes and program consistency phases of performance assessment tasks evaluated after their implementation by students, instructors and administrators? .....	135
5.2.3. What were the stakeholders' views on the continuity of the current performance assessment implementation? .....	138
5.3. Conclusion.....	141
5.4. Implications.....	143
5.5. Suggestions for further studies.....	145
<b>REFERENCES .....</b>	<b>147</b>
<b>APPENDIX A.....</b>	<b>156</b>
<b>APPENDIX B .....</b>	<b>158</b>
<b>APPENDIX C .....</b>	<b>160</b>
<b>APPENDIX D.....</b>	<b>161</b>
<b>APPENDIX E .....</b>	<b>166</b>
<b>APPENDIX F .....</b>	<b>170</b>

<b>APPENDIX G.....</b>	<b>171</b>
<b>APPENDIX H.....</b>	<b>172</b>
<b>APPENDIX I.....</b>	<b>173</b>
<b>APPENDIX J.....</b>	<b>176</b>
<b>APPENDIX K.....</b>	<b>196</b>



## LIST OF TABLES

Table 1	The EFL Instructor and student distribution by language levels	51
Table 2	The assessment percentages in 2016 Fall Semester before PTs	53
Table 3	The assessment percentages in 2017 Spring Semester (PT Trial)	54
Table 4	The assessment percentages in 2017-2018 (Current PT implementation)	55
Table 5	Weekly Schedule of Assessment at Preparatory School	62
Table 6	Cronbach's Alpha information for instructors' scale	73
Table 7	Cronbach's Alpha information for students' scale	73
Table 8	Positive Aspects of PTs as expected by students before implementation	79
Table 9	Negative Aspects of PTs as expected by students before the implementation	81
Table 10	Positive Aspect of PTs as expected by instructors before the implementation	84
Table 11	Negative Aspects of PTs as expected by instructors before the implementation	86
Table 12	Positive Aspects of PTs as expected by administrators before the implementation	88
Table 13	Negative Aspects of PTs as expected by administrators before the implementation	89
Table 14	Total mean values of students' opinions on the phases of PT implementation	90
Table 15	Opinions of students on the planning phase after the PT implementation	91

Table 16	Opinions of students on the application phase after the PT implementation	92
Table 17	Opinions of students on the scoring phase after the PT implementation	93
Table 18	Opinions of students on the learning outcomes phase after the PT implementation	94
Table 19	Opinions of students on the program consistency phase after the PT implementation	95
Table 20	Total mean values of instructors' opinions on the phases of PT implementation	96
Table 21	Opinions of instructors on the planning phase after the PT implementation	97
Table 22	Opinions of instructors on the application phase after the PT implementation	98
Table 23	Opinions of instructors on the scoring phase after the PT implementation	99
Table 24	Opinions of instructors on the learning outcomes phase after the PT implementation	100
Table 25	Opinions of instructors on the program consistency phase after the PT implementation	101
Table 26	Total mean values of instructors' opinions on the phases of PT implementation	102
Table 27	Opinions of administrators on the planning phase after the PT implementation	103
Table 28	Opinions of administrators on the application phase after the PT implementation	104
Table 29	Opinions of administrators on the scoring phase after the PT implementation	105



Table 30	Opinions of administrators on the learning outcomes phase after the PT implementation	106
Table 31	Opinions of administrators on the program consistency phase after the PT implementation	107
Table 32	Reasons of students for the continuity of PTs in the same way	110
Table 33	Suggestions of students for the PT implementation in the future	112
Table 34	Reasons of students for not continuing the PT implementation	114
Table 35	Reasons of instructors for the continuity of PTs in the same way	116
Table 36	Suggestions of instructors for the PT implementation in the future	118
Table 37	Reasons of instructors for not continuing the PT implementation	120
Table 38	Suggestions of administrators for the PT implementation in the future	121

## ABBREVIATIONS

BTS	Barlett's test
CLT	Communicative Language Teaching
CSL	Computer Supported Learning
EFL	English as a Foreign Language
ELT	English Language Teaching
ER	Extensive Reading
ESL	English as a Second Language
FLED	Foreign Language Education Department
HEC	Higher Education Council
KMO	Kaiser-Mayer-Olkin
PBA	Performance-based Assessment
PPP	Power Point Presentation
PT	Performance-based Assessment Task
TA	Traditional Assessment
SD	Standard Deviation
SS	Students
T	Teacher
YL	Young Learner

## SIGNIFICANT TERMS

- Alternative Assessment:** a means of assessing learner performance via a variety of open-ended, communicative, or creative techniques (Brown, 2004).
- Assessment:** the various means of accumulating information about a student's development and success including informal practices administered by the teacher during the course (Thornbury, 2006).
- Evaluation:** a process that is conducted to gain data in order to make alterations and modifications, eliminations and/or accept something in the curriculum (Ornstein & Hunkins, 1998).
- Performance-based Assessment:** 'performance-and-product assessment' since performance assessment may contain both performance and product of students (Fitzpatrick & Morrison, 1971).
- Program Evaluation:** a systematic operation in constant change including data collection, observations, and analyses that concludes in a value judgement related to the program being evaluated, or one or more of its components (Mızıkacı, 2006).
- Traditional Assessment:** a test that is identified as a means of testing learner achievement by administering objective questions which contain one correct or best answer (Brown, 2004).

## CHAPTER 1 INTRODUCTION

### 1.1. Background of the Study

In the contemporary educational field, the demands for evaluation are rising. Language teaching institutions have been seeking more accurate, convenient and practical solutions to design and carry out their language programs in accordance with their educational goals. As the last stage of the cycle, program evaluation plays a significant role on their decisions. According to Norris (2009), in this era there have been robust demands for accountability testing, institutional accreditation, outcomes assessment, and quality control so language educators are developing a heightened awareness of program evaluation and its prominent role in determining how language teaching and learning occurs. The contribution of evaluation on understanding and improving language teaching practices and language programs is undeniable.

Among various components, assessment takes a crucial part in evaluation. An effective assessment must target the collection, analysis, and interpretation of measures of student learning and performance. The data collected from student outcomes must serve in measuring the effectiveness of each academic program and of an entire department (Gaies, 1992). The novice developments in every field of English language teaching and hence the field of testing are to be followed.

Considering these innovations, today there has been a shift from the traditional assessment methods to the alternative ones in the field of assessment. However, institutions mostly have some concerns about alternative assessment in terms of grading, validity and reliability aspects. Some practitioners argue that using alternative assessment methods such as performance assessment tasks (PTs) may not

be an effective and practical way whereas some others suggest that this implementation is worth trying despite its negative aspects.

This study aims to evaluate the whole implementation process of performance assessment tasks in its first semester in an EFL preparatory school as a supplementary assessment practice to traditional assessment practices in the current context and its effectiveness from the stakeholders' perspective. PTs are claimed to ensure measuring abilities such as communicative ability, problem-solving, and critical thinking that are troublesome or impossible to measure with standard tests by Perlman (2003). In this respect, the effectiveness of the combination of both performance-based assessment (PBA) and traditional assessment (TA) will be investigated. The main actors of teaching and learning cycle, teachers- learners-administrators, play a significant role on the evaluation process. According to the implications of this research, the PT practices will be interpreted within the academic program it serves for. It is planned to put the necessary revisions and changes suggested by the stakeholders into practice for ongoing improvement. Thus this study based on scientific evidence may lead to more convincing arguments in the fields of both assessment and evaluation.

## **1.2. Statement of the Problem**

Evaluation is a process carried out to determine whether to make changes, to make modifications, eliminations and/or accept something in the curriculum (Ornstein & Hunkins, 1998). Assessment, which is an inseparable element of teaching and learning, has been a critical issue in terms of not only language teaching practices but also program evaluation procedures for quite a long time. This significant component requires some essential changes in the light of innovative developments in ELT. On the contrary to new trends in assessment, in language classes, it is often observed that the traditional assessment methods have still been executed predominantly as a testing

tool in many schools around the world, particularly in Turkey where this study is conducted. They are mainly used to assess language learners' overall success in language skills and areas. These are favored by practitioners due to their standard way of testing and in terms of practicality, reliability and validity issues (Nasab, 2015). Yet, traditional assessment (TA) in the EFL context is still a contradictory issue. It may be criticized by practitioners in several aspects such as its weakness in measuring students' performance and overall learning process. For instance, Bailey (1998) criticizes traditional assessment as being indirect and inauthentic. Law & Eckes (1995) argue that traditional tests cannot present sufficient information about a student's progress or the challenges he/she may have faced during the test.

Lately in testing trends there has been a shift from traditional assessment methods to alternative ones. There seem to be many positive aspects of these practices. Wiggins (1990) states that through authentic assessment, students benefit from a large number of language learning skills and they can reflect their gained ability as their performance on the tasks. Performance tasks (PTs), presented as a way of alternative assessment method in this study, are defined as either the observation of behavior in the real world or a simulation of a real life activity (Weigle, 2002). Creating assessment tasks effectively requires some stages such as pondering upon curriculum content related to learning outcomes, designing performance tasks to demonstrate students' achievement of these outcomes and generating criteria for the assessment (Cohen, 1995). In addition to its challenging creation process, in EFL setting, where authenticity is a necessity but may rather be a difficult concept to construct, it is a problematic issue to implement alternative assessment methods such as performance-based assessment (PBA) and integrate it into traditional ways of assessment effectively in this context. For instance practitioners have concerns about

the accurate implementation of alternative assessment to large-scale classes (Worthen, 1993). Unfortunately this is an undeniable situation for most of the preparatory classes at state universities in Turkey due to excessive student population.

The present study aims to evaluate performance-based assessment tasks integrated into the assessment practices as a supplementary component of the language program in a preparatory school in Turkey, where the assessment methods are based on predominantly TA practices. The performance-based assessment implementation targets to help learners transfer their competence to performance by its nature. As language learning is a long process, it is aimed to focus not only on the product but also the process. However, Turkish students get used to a TA background, which is based on rote-learning and only product-oriented assessment. The number of the paper-based tests is criticized by the practitioners and learners for different reasons such as stealing from the class-time, focusing more on competence, not being learner-centered, disregarding learner autonomy and failing to measure students' actual performance in the target language. Although some process-oriented attempts have been carried out in the target institution, the student performance has been assessed by TA methods. In order to measure learners' language knowledge, particularly their productive skills, more performance-based and process-oriented alternative assessment techniques are to be executed. Using merely traditional assessment methods is considered as insufficient to encourage students' production in the target language and to assess it. The novice practices should be planned to serve for all language skills and language elements. As a remedy for all these problems, performance-based tasks as a supplementary tool are combined with TA practices in the target institution whose program is under investigation of this study. The prospective impacts of the combination of TA and PBA on stakeholders' opinions

regarding its consistency with the current language program are evaluated based on the questions addressed by Decision/Accountability-Oriented evaluation model of Stufflebeam (2001).

### **1.3. Significance of the Study and Research Questions**

From the assessment perspective, this study aims to investigate an alternative assessment practice as consisting of a combination of a traditional and alternative assessment in detail by presenting the possible stages of its implementation and its probable contribution to language learning of Turkish adult students. Various ways of using PTs and their possible reflections should be taken into consideration while integrating these applications into curricula in order to make them more efficient for both teachers and students. Evaluating the data obtained from the stakeholders is considered significant so as to form a broader overview. The study mainly focuses on both positive and negative aspects of PTs and several phases of its implementation in the EFL context from the perspective of stakeholders. Firstly, it examines the expectations of the stakeholders on these PT practices. Secondly, it investigates the stakeholders' opinions of the overall process including planning, application, scoring, learning outcomes and its place in the current English language program. Thirdly, it evaluates the opinions and suggestions of the stakeholders on the continuity of the PT implementation in the following years. Finally the whole PT process is discussed in the light of decision-making/accountability-oriented evaluation principles as a part of improvement-oriented evaluation approach. Moreover, the study presents possible applications of PTs in several language skills and areas and in three different language proficiency levels. Therefore this study may enable the language teaching institutions to make sound judgments in their decision phase.



Research on performance-based assessment mostly targeted and conducted with young learners (Gönül, 2010; Paksoy, 2015; Pulat, 2014). Different from the other previous studies in ELT, the target group of this mixed-methods study is adult learners. Through this research, light can be shed on institutions which teach foreign languages to adult learners.

From an institutional aspect, the main goal of the study is to probe the various aspects of performance-based assessment (PBA) as a way of measuring actual student performance and an assessment tool to increase learner autonomy and motivation, as claimed in the literature. The findings can be utilized to evaluate and determine the assessment practices suitable for the target institution as well as similar contexts. Therefore the suggestions and opinions deduced from the study may shape and enhance the assessment component of the current English language program as proposed in “Improvement/Accountability-Oriented Evaluation Approaches” by Stufflebeam (2001, p.42).

From a broader evaluation perspective, this study employs an improvement-oriented formative evaluation of the PT practices through the stakeholders’ perceptions before the implementation and their opinions after the implementation. Thus the findings and implications might be used for the future program evaluation decisions by other institutions in EFL contexts. Despite the fact that there have been a number of studies on overall evaluation of educational programs, a detailed analysis of a specific component of a program is still quite rare. This study may serve as a sound example of this type of research in the literature.

In this study, answers to the following research questions are explored:

1. What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of
  - a. students?
  - b. instructors?
  - c. administrators?
2. How were the planning, application, scoring, learning outcomes and program consistency phases of performance assessment tasks evaluated after the PT implementation by
  - a. students?
  - b. instructors?
  - c. administrators?
3. What were the views on the continuity of the current performance assessment implementation as stated by
  - a. students?
  - b. instructors?
  - c. administrators

#### **1.4. Overview of the Methodology**

The first research question was aimed to present the general view of stakeholders' expectations on PTs before the actual implementation. In order to answer this research question, an open-ended questionnaire was used as an instrument to collect data from randomly selected 126 students from three different English proficiency level of a preparatory school at a state university in Turkey. By adding two classes per level, in total six classes were used. The data collected was analyzed through content and thematic analysis assisted with the expert opinion.

The second research question was aimed to evaluate the phases of the PT implementation and its consistency with the current language program through three 5-point Likert-scales in order to gather the opinions of the stakeholders after the implementation. The data collected from 129 of students from the same classes used to collect the preliminary qualitative data. The data was analyzed through a statistical program. The reliability and validity of these scales were checked via the statistical program under the control of experts.

The last research question was aimed to present the stakeholders' opinions on the continuity of the current PT implementation. As an instrument, the qualitative section of the scale was utilized. Three open-ended questions were asked to the stakeholders. The qualitative data was analyzed through the content and thematic analysis under the control of the expert.

#### **1.5. Limitations of the study**

Since the population and the sample numbers ( $N/n = 2$ ) for administrators were the same and the number was limited to two, the piloting procedure was eliminated and also the reliability and validity of that scale could not be checked.

All the data collected through the questionnaires were based on the opinions of the participant students, instructors and administrators. Their expectations and opinions on the PT implementation under the investigation of this study were assumed to be genuine.

This research was a small scale-study findings of which were considered limited to the stakeholders at a certain preparatory school of a state university in Turkey. However, the instruments and the PT implementation procedures may be adapted to various language learning settings as mentioned in the implications section.

As this study involves formative evaluation of a 15-week application, PT implementation in the long run and its impact should be investigated to make sound judgements.

## CHAPTER 2 LITERATURE REVIEW

This chapter reviews the literature on program evaluation, program evaluation types and models, a specific program evaluation approach, and research on program evaluation. Additionally assessment in English language teaching, traditional and alternative assessment, studies on the use of various alternative assessment methods, performance assessment tasks as an alternative assessment method, research on performance assessment tasks by displaying the teacher, learner and administrator perceptions and opinions are presented.

### **2.1. Program Evaluation**

Program evaluation has recently taken a crucial place particularly in terms of language education due to the endurance to keep pace with its constantly changing and evolving nature. However before investigating the research on program evaluation, the concept of evaluation, as a broader term, is to be taken into consideration. Evaluation is defined as a study that is designed and carried out to assist some audience to evaluate an object's merit and worth by Stufflebeam (2001), who has also identified twenty-two models of evaluation in the same article. Considering evaluation as a means of innovation, revision and modification in language programs, it is best identified as a process. Therefore as in Ornstein & Hunkins's definition (1998) evaluation is defined as a process that is conducted to gain data in order to make alterations and modifications, eliminations and/or accept something in the curriculum. Evaluation is to be carried out systematically in order to reach its predetermined goal. It is essential to observe that an evaluation allows various aims, and different stakeholders might ponder upon these purposes. The evaluators are responsible for negotiating clear and operable purposes with the program stakeholders

thus it can be carried out with proper methods and can be benefitted (Yang, 2007).

Considering language education field, evaluation is an inseparable fragment of language learning-teaching contrary to what previously conceived, and a process that has the potential to illuminate and ensure the value of language education (Norris, 2009).

Program evaluation, which is the main focus of this study, can be described as a systematic operation in constant change including data collection, observations, and analyses that concludes in a value judgement related to the program being evaluated, or one or more of its components (Mızıkacı, 2006). In order to keep pace with the constant alterations in the developing and evolving world, curricula and teaching programs also evolve. Hence these programs need to go through a continuous cycle of planning, application and evaluation (Topkaya & Küçük, 2010).

### **2.1.1. Types and Models of Program Evaluation**

In terms of evaluating a program, there are not only different models, but also various types of it. Evaluation can be categorized as formative and summative. According to Yang (2007) formative evaluation serves the goal of enhancing a program by investigating program practices and supply feedback on them. Summative evaluation provides educators to make judgments about a program by deducing the effectiveness of overall program, usually influencing the fate of the program. Long (1984) stresses the three differences between formative and summative evaluation as focus, timing and purpose. Program evaluations can also be categorized as internal and external in terms of its purpose. Gaies (1992) states that every systematic program evaluation has two different objectives such as internal and external. Despite the fact that various types and models of program evaluation are administered in different

educational institutions, the main issue is to determine the most effective one for the context in which the evaluation is administered.

In this study, in accordance with previous research, mainly formative and internal evaluation through improvement-oriented approach is adopted. While planning the evaluation, the appropriate design for the study is shaped by the questions to be answered in order to provide a cohesive and convincing evaluation (Weiss, 1998). Due to the questions addressed in the “Decision/Accountability-Oriented Approach” presented by Stufflebeam (2001, p. 42) under the name of “Improvement/Accountability-Oriented Evaluation Approaches”, this evaluation approach seems to be consistent with the purpose and nature of this study. In the cases of the implementation of decision/accountability-oriented approach, both formative and summative evaluations are required by the program staff and other stakeholders. Besides it can enable both internal and external evaluation (Stufflebeam, 2001, p. 58). Although this research study mainly focuses on formative evaluation, the outcomes of the study may direct the evaluator to the point of a summative evaluation in which final decisions are made about the program. In the light of decision/accountability-oriented approach, the questions addressed by this approach are aimed to be answered in the present study.

### **2.1.2. Decision/Accountability-Oriented Approach**

Among twenty-two evaluation approaches discussed in the article by Stufflebeam, the decision/accountability-oriented approach is presented as one of the best and most applicable approaches by the researcher (2001, p. 42-58). This approach is presented as a part of Improvement/Accountability-Oriented Approaches. There are two others, namely consumer-orientation and accreditation/certification approaches labeled under the same title. According to Stufflebeam (2001), the

common points of these are mainly seeking all relevant outcomes, being objectivist, utilizing both qualitative and quantitative methods, considering stakeholders' needs as the essential criteria and being improvement-oriented.

Stufflebeam (2001) highlights the improvement aspect of this evaluation approach. The decision/accountability-oriented approach differs from others as aiming to enhance the program under the investigation as well as judging its merit and worth. It attempts to find answers for context-limited questions by integrating all relevant stakeholders into the evaluation process. As it can be inferred from its name, it assists the decision making process. Accountable evaluative information including assessment of program materials and facilities and assessment of short-range or long-range outcomes can be provided by means of this approach. It emphasizes not proving but improving as the key goal of the program evaluation. Various methods may be utilized such as surveys, case studies, observations, or interviews in order to collect feedback from stakeholders. The evaluator should interact with the stakeholders and inform them. Sharing the interim reports on the evaluation with the stakeholders may be useful to engage them in the process and thus to improve the program in cooperation. This approach is valid for both formative and summative evaluation as well as proper for internal and external evaluation. Considering the weak points of the decision/accountability-oriented approach, Stufflebeam (2001) pinpoints that the evaluator should ensure his/her objectivity and the summative evaluation should be provided properly. This approach is claimed to emphasize the formative evaluation more than the summative one.

As the main focus of this approach is improvement by integrating the relevant stakeholders into the change process and by putting more emphasis on the formative evaluation, the decision/accountability-oriented evaluation approach is selected



suitable while evaluating the assessment component of the current language program in this study. Moreover the questions addressed by this approach are also found related with the objective of the study by the researcher. The questions presented by Stufflebeam, (2001) can be listed as:

“Has an appropriate beneficiary population been determined? What beneficiary needs should be addressed? What are the available alternative ways to address these needs, and what are their comparative merits and costs? Are plans of services and participation sound? Is there adequate provision for facilities, materials, and equipment? Is the program staff sufficiently qualified and credible? Have appropriate roles been assigned to the different participants? Are the participants effectively carrying out their assignments? Is the program working and should it be revised in any way? Is the program effectively reaching all the targeted beneficiaries? Is the program meeting the participants’ needs? Did beneficiaries play their part? Is the program better than competing alternatives? Is it affordable? Is it sustainable? Is it transportable? Is the program worth the required initial investment (p. 56-57)?”

### **2.1.3. Research on Program Evaluation**

There have been various research studies conducted in the field of program evaluation. They vary in terms of the applied program models, program designs, methodology, context and naturally their findings. For instance in a longitudinal case study conducted by Gunn (1999) evaluation of computer supported learning (CSL) has been investigated through program improvement approach in terms of its educational effectiveness and academic credibility. According to its findings the evaluation framework used in the study confirms that it was useful and appropriate for its purpose. The study fulfilled by Yoon-Hee, Byung-Kyoo & Hyung-Sun (2008) in a different context targeted to provide a holistic evaluation of an in-service teacher training program by assessing proficiency gains as well as tracing on-going development of teaching skills through process-oriented procedures. Another evaluation study administered by Yang (2007), whose participants were both administrators and teachers, produced quite useful findings which were used to

enhance the program and its evaluation process was also useful in providing positive program changes via utilization-focused evaluation model.

Compared to the general history of program evaluation, which is a crucial component of program development by presenting the strengths and weaknesses of the relevant program, in many other countries program evaluation is considered as a separate field of science requiring expertise. However, regarding the program evaluation in Turkey and the attempts to improve the program according to the study findings, there seems to be no consistency (Çeliker, 2015). For the sake of evaluating the effectiveness of a language program, some useful studies have recently been conducted in this field for Turkish educational arena. For instance, a study conducted by Daloglu & Coşkun (2010), which drew attention to the importance of program evaluation for teacher education programs, focuses on the pre-service English teacher education program components in a Turkish university context by using Peacock's (2009) recent evaluation model. That study was considered effective in providing multidimensional feedback about the program being evaluated. The study results revealed that the pedagogic side of the program needed to be enhanced. In his study, Mammadov (2015) investigated the existing policies on the education of gifted K-8 students and conducted a program evaluation focusing on the strengths and weaknesses of the gifted programs in public and private schools in order to address the effectiveness of those programs. In a qualitative research by Yavuz & Topkaya (2013), they examined the perceptions of teacher educators regarding the changes in the English Language Teacher Education Program introduced by the Turkish Higher Education Council (HEC) in 2006. The data analysed yielded that while teacher educators found some of the changes appropriate, they raised far more serious concerns with the new program. In a study conducted by Çeliker (2015), which aimed

to determine the program evaluation self-efficacy levels of Educational Sciences and Teacher Training Experts and to examine these self-efficacy levels according to various variables, findings indicated that the program evaluation self-efficacy level of the relevant experts was satisfying. Çeliker (2015) claims that through the program evaluation practices, feedback was taken about the program and in accordance with this feedback some essential decisions which affect the quality of education are made such as continuing with the program, revising it or eliminating it completely. In another research conducted by using impact evaluation, one of Owen's (2007) five forms of evaluative inquiry, Celen (2016) evaluated the practicum program offered at the Foreign Language Education Department (FLED) at an English-medium state university in Turkey. Celen states that the evaluative model undertaken in that study indicated similarities with the "improvement-focused approach" used in Gunn's research in 1999.

Even though there have been attempts of sound research in program evaluation in Turkey as mentioned above, it can be stated that this concept should be regarded more in educational field in Turkey. Despite the fact that it is a relatively new concept in language teaching, program evaluation is a promising field in the sense of its impact on every aspect of a language program. Weiss claims that considering the contribution of evaluation findings, they often have a crucial effect by providing new concepts, perspectives, new means of making inferences and potential directions (Weiss, 1998 as cited in Yang, 2007). As in this study, one component of the program can be evaluated and the results can provide an insight for not only the future renewal attempts in the relevant component investigated but also the overall entity of the language program as every single part is connected firmly to that whole.

## **2.2. Assessment**

### **2.2.1. Assessment, Evaluation and Testing**

The terms “assessment”, “evaluation” and “testing” are used interchangeably and there is a common tendency for misunderstanding or misusing these in educational field although they refer to distinct concepts (Can, 2017; Önal, 2010). According to Önal, evaluation is interested in all aspects that form and influence the whole learning and teaching process in terms of its components such as course design, the objectives of a syllabus, and materials used whereas assessing is related to the performance of the student measured by the teacher. Through assessment, information is collected in several ways on the learner’s language skill and success (Önal, 2010). To stress the distinction of these two terms, Can (2017) defines assessment, which is a comprehensive study field, as an ongoing process and even though it is closely related to the terms such as testing and evaluation, it is a different concept. In his book which consists of clear definitions for the ambiguous methodological terminology, Thornbury (2006, p. 18), who also highlights the distinction among these three concepts, defines assessment as the various means of accumulating information about a student’s development and success including informal practices administered by the teacher during the course. Brown (2004), who also identifies assessment as a process, emphasizes the distinction between assessment and testing. He claims that assessment, as a broader concept, is an ongoing process which occurs in various occasions such as a student’s response to a question, a student’s offer to a comment, or his/her trial on a new lexical item or form, the teacher subconsciously monitors the student’s performance to make an assessment of it. However, testing can be defined as administrative procedures occurring at specific times when the students are aware that their performance is being tested (Brown, 2004). In terms of the program

evaluation, assessment, a systematic collection of information to improve students' learning and development, is a crucial process of a program in which learners are tested whether they are learning or not. Through various kind of measures, it is checked and evaluated if students' learning goals determined by their faculties for their courses have been attained or not (Gürsoy, 2013). As a constant data collection process, assessment elicits the relationship between educational programs and student learning and performance. The data collected via assessment triggers evaluation processes leading to judgments about effectiveness and quality issues. Assessment means more than merely giving grades or marks, and it is about determining the weaknesses and strengths, making judgments and resolving on the right and the wrong and the good and the bad (Göçtü, 2013). In that sense, assessment has a significant role in the education process as one of the most important parts of the learning and teaching process, both for students and practitioners. Assessment affects teachers' decisions on when, what and how to teach and it gives students the opportunity to monitor what they have learnt by raising their awareness (Zaimoglu, 2013). Every curriculum requires different types of assessment practices, so utilizing only one type of assessment is not adequate. Considering its great impact on teachers and students, it is a valid concern to determine the suitable assessment tools serving best for the instruction (Gökçen, 2005).

### **2.2.2. Assessment Types**

The assessment types can be categorized in various ways. The first category can be named as formative and summative assessment. While formative assessment is concerned with gaining mastery of a subject, summative assessment with a purpose to grade and sort students focuses on performance (Carless, 2011). From a broader perspective, formative assessment, on-going or dynamic assessment, is also known as

‘assessment for learning’ whereas summative assessment, administered periodically, like at the end of a year or a term or a unit, is also known as ‘assessment of learning’ (Büyükkaracı, 2010). Another distinction can be made between formal and informal assessment. Formal assessments are systematic and planned sampling techniques like exercises or procedures particularly prepared to tap into skills and knowledge. On the other hand, informal assessment forms can range from unplanned, incidental comments and responses to coaching and other impromptu feedback to the student (Brown, 2004). In terms of educational evaluation, assessment consists of two major types: traditional testing and alternative assessment (Huerta- Macias, 1995), which are in the centre of discussion in this study.

### **2.2.3 Language Assessment & Trends in Language Assessment**

Language assessment, the significant part of the present study, plays a central role in applied linguistics by putting its theories into operation and providing its researchers with data for their studies on language knowledge or use. Considering the recent history of language assessment, in the late 1970s with the advent of communicative teaching, testers were supposed to plan new theories of language testing (Clapham, 2000). For language testing, communicative testing, which attempts to measure a much broader range of language abilities, was the dominant trend in the 1980s. The major impetus from language teaching leads to many innovations in language testing over the years. Due to these advancements in the 1980s, language testing has emanated as a discipline in its own right within applied linguistics. Based on the reviews in the 1990s, they all reflected optimistic views of where language testing was going and what it had to offer other areas of applied linguistics (Bachman, 1991). Alderson (1991) claims that developments in the content of language tests have been made to meet the requirements of communicative performance tests

(Alderson, 1991 as cited in Farhady, 2005). Within the last few decades, language assessment has witnessed a multidimensional progress. The literature in assessment poses that major developments in language assessment have occurred in four areas ,namely forming theoretical models of language ability, facilitating sophisticated statistical techniques in test data analysis, designing content-specific language tests and exploiting computer technology not only to administer and score but also to analyse the test data. These advancements in each field has resulted in success independently from one another offering a comprehensive treatment of language assessment (Farhady, 2005). Even advances in technology have led to alterations in language assessment such as computer-based testing, which has achieved a rapid growth in the past decade. Now they are used to deliver language tests in many contexts at all stages in the test development and administration process (Alderson & Lancaster, 2001). In accordance with its historical evolvment, in education around the world, language assessment has turned out to be a major policy issue in recent decades. It has raised a large number of fundamental issues about the conceptualizations, functions, and implementations of assessment related to teaching, learning and curricula. Educational programs adopt curriculum standards defining explicit criteria for language learners' expected language achievements and multiple challenges they face (Cumming, 2009).

Owing to the growing trends in language teaching-learning focusing more on communication, highlighting performance rather than competence and adopting more learner-centred methods, the traditional teaching methods and their testing components have been insufficient to meet the requirements of these developments. As a result, the field of assessment has shifted to a new paradigm: the assessment paradigm (Farhady, 2003). In pursuance with these developments, assessment has

witnessed a paradigm shift from the traditional assessment to the alternative assessment. (Bayram, 2015; Canagarajah, 2006; Farhady, 2003; Özeren, 2013; Richards & Renandya, 2002). This paradigm shift, which has led to a great number of changes in assessment procedures and practices, is discussed in a more detailed way in the following sections of this literature review.

### **2.3. Language Assessment at Higher Education (Preparatory schools) in Turkey**

Many universities in Turkey consist of preparatory classes mostly obligatory for the students which have Foreign Language Programs, predominantly in English. A placement or proficiency test is administered to preparatory school students in the beginning of the academic year. Gönen (2013) explains the preparatory school process as students having to study for a period of one year to three terms in some universities to reach an adequate level of English in intensive programs whose weekly-hours range from 20-30 hours. If the students cannot meet the requirements of the programs' goals within two years, they are not allowed to continue their higher education in English taught programs (Gönen, 2013). There is a fail and pass or in other words a matter of life and death for these students, which causes a great pressure on both the instructors and students. Therefore, the administration of the preparatory schools constantly encourages the practitioners to use applicable teaching methods in order to achieve success in English teaching at Higher Education English Language Programs. However, they tend to avoid coping with the assessment issue. Mostly teachers are expected to obtain data on the progress of their students' learning as assessment activities (Gönen, 2013). Doğan (2009) claims that assessment practices carried out at higher education institutions focus predominantly on determining the level of proficiency as summative assessment and measuring students' achievement through assessments which are supposed to be completed within a given time. He also



states that the common types of assessment administered at preparatory classes are classroom assessment such as multiple-choice, short-answer, true-false tests, most of which are effective in assessing the content-related language knowledge while they fail to measure writing skills, critical thinking or creativity (Doğan, 2009). The traditional tendency in teaching and assessment is usually dominant at higher education context. In educational field in Turkey, it is more common to see teacher-centred learning settings rather than student-centred ones (Sönmez, 2013).

Considering general alteration in ELT, in the past fifteen years with the intention to direct attention to student learning from many different aspects and particularly with the impact of authentic and alternative assessment types, assessment has gradually become significant in higher education (Göçtü, 2013). By adopting a new innovative perspective in education, more student-centred practices have been carried out, which leads to a more student-centred approach in assessment practices (Özeren, 2013). In the evolving education milieu, assessment has gained different meanings and standards for learners. Thus these changes trigger the shift from traditional assessment to alternative assessment (Bayram, 2015).

In accordance with these developments in ELT, correspondingly in assessment there have been some attempts for alterations at higher education institutions particularly at preparatory schools in Turkey. The increasing number of the research on alternative assessment indicates this shift. Today alternative assessment implementations are becoming more and more common among the distinguished universities of Turkey.

#### **2.4. Traditional Assessment**

Assessment can be categorized in various ways, some of which are abovementioned, such as formal and informal assessment, formative versus

summative assessment, norm-referenced versus criterion referenced tests, discrete-point versus integrative testing, high-stakes versus low-stakes tests, direct and indirect testing, objective versus subjective testing and traditional versus alternative assessment (Taş, 2016). Among all these classification systems developed to assess student learning, one way of grouping these is putting them into two major categories: traditional and alternative assessments (Zaimoğlu, 2013). Traditional assessment, typically a summative test, is identified as a means of testing learner achievement by administering objective questions which contain one correct or best answer (Brown, 2004). Traditional assessments refer to examples such as multiple choice, matching, true/false, fill-in-the-blank, short answer, and essay exams. Due to their long use in education, these types of assessments are named after the term ‘traditional’ (Streff, 2016).

#### **2.4.1. Pros of Traditional Assessment**

Even though most educators, being aware of negative aspects of TA, believe in the necessity of using a variety of alternative or authentic assessment techniques, traditional testing methods are still the most commonly used assessment tools in language classrooms. This may be due to the lack of knowledge or experience with alternative assessment methods (Gökçen, 2005). Can (2017) also claims that using paper-pencil tests involving a wide variety of procedures such as choosing the appropriate assessment methods, administering quizzes, writing test items or revising any teacher-made test to match the instructional objectives is still the most commonly used classroom assessment practice. There have been some sound reasons for traditional assessment to remain superior to the alternative assessment methods. For instance, in a meta-analysis research conducted by Özeren (2013) it was seen that there was especially a big concern about the application of the new assessment and

evaluation materials in terms of the lack of time, which has been eliminated in the following years, and classroom sizes. Yüce (2015) claims that the reason why practitioners mostly prefer to administer traditional assessment rather than alternative assessment to measure their students' performance is that they are generally used to administering TA methods, thus feel quite confident. Büyükduman (n.d), who concluded a study on alternative assessment at a prep school in Turkey with positive results towards alternative assessment, regarded alternative assessment methods as a challenge for not only the teachers but also the students who had been exposed to a different assessment system, typically based on rote learning, during their whole prior educational period. In another study (Al-Nouh, Taqi & Abdul-Kareem, 2014) conducted with primary school teachers, the findings have displayed that alternative assessment is time-consuming and it ignores pupils' writing skills. They supported the traditional written tests instead. Another research administered to the students in Belgium, they favour multiple-choice format exams to essay type questions (Struyven, Dochy & Janssens, 2005). Considering policy decisions from a broader perspective even though innovative, criterion-referenced assessments meet the pedagogical requirements within educational systems, the authority of norm-referenced language tests preserves its place in educational field (Cumming, 2009). 'Standardised test' is typically a norm-referenced test defined as a test whose difficulty level is known. This type of test has been adequately piloted and analysed, whose results can be compared with those of a norming population (Alderson & Banerjee, 2001). As the representatives of traditional assessment, they are still mainly favourable in terms of reliability and validity, which is a common concern for alternative assessment techniques. Bachman & Palmer (1996) mentioned that certain test qualities are essential for a test to achieve its usefulness (as cited in Taylor, 2006).

Taylor stressed four dimensions of assessment in accordance with these test qualities, namely validity, reliability, practicality, and impact (2006). By the opponents of alternative assessment, these practices have been criticized in terms of disregarding validity and reliability and also their implementations have been regarded as not practical. As Gürsoy (2013) stated, valid and reliable test scores provide fair assessments by producing meaningful results by eliminating bias and preventing unfair advantages through testing the same or similar information under the same assessment conditions.

#### **2.4.2. Cons of Traditional Assessment**

Considering the basic characteristics of traditional ways of testing, from a more innovative perspective, they can represent merely a fraction of what learners want to produce, they play not a developmental but a judgemental role, they often fail to improve the overall quality of language learning and teaching, they are summative and usually developed and administered by outsiders, they cover a great amount of teaching time, they are more interested in the discriminative numerical marks, useless descriptions than the formative feedback, they tend to treat learners as passive participants of the learning process, they are teacher-centred, they adopt limited time procedures giving learners one chance to display competence, they are anxiety-generating, and may also be demotivating, they are administered to large groups of students and in this respect they disregard individual differences (Toumi Lafi, n.d.). All these characteristics are also their essential reasons of being in the centre of criticism by authorities recently. According to Richards & Renandya (2002), old paradigm in teaching and assessment focuses on language and it is teacher-centred. It also adopts skills isolated, emphasizes on product with only one answer and one-way correctness approach by administering tests for the sake of testing. After more

innovative teaching and assessment paradigm has appeared, as Brown (2001-2003) claims TA practices are insufficient to determine multidimensional aspects of students' learning skills and they are ineffective to assess student performance and particularly productive skills. Hence these are mostly preferred to measure student achievement in receptive skills. Conventional assessment does not give students an opportunity to construct their own knowledge and meaning about what they have learnt (Zaimoğlu, 2013). Traditional assessment usually fails to assess students' communicative skills and do not encourage their involvement in their own evaluation. Alternative assessment tools, which are more student-centred emphasizing integrated skills and the process of learning, play a significant role to promote student engagement in the assessment and evaluation process (Büyükduman, n.d.). Traditional assessment techniques fall behind current EFL teaching methods increasing student motivation, encouraging student learning and providing a clear understanding of student achievement because TA techniques are not enough to assess multiple dimensions of language learning and fail to assess the full range of student outcomes (Burnaz, 2011).

According to the recent studies and reviews in the literature, traditional assessment is losing its popularity whereas alternative assessment techniques are increasing in popularity. In a study conducted by Chan (2008) on EFL elementary school teachers' beliefs and practices of multiple assessment, none of the teachers have chosen traditional tests as the primary assessment. Another study by Cirit (2015) investigated perceptions toward traditional, online and alternative assessment. According to his findings, in general, ELT pre-service teachers preferred alternative assessment to online or traditional assessment. In a prep-school in Turkey, a similar context to the one in this study, according to the study results deduced by Burnaz,

(2011), most of the students preferred portfolio assessment to traditional assessment due to the defects of TA such as time pressure, leading to memorization and not measuring English speaking skills effectively. Moreover, participants claimed that they were dissatisfied with TA practices and complained about traditional assessment system.

## **2.5. Alternative Assessment**

Alternative assessment has been defined as an alternative to standardized testing and all of the problems related with such testing. No single definition of alternative assessment is found, yet a variety of labels have been presented to distinguish it from traditional, standardized testing (Huerta-Macías, 2002). According to Brown's definition (2004) alternative assessment is a means of assessing learner performance via a variety of open-ended, communicative, or creative techniques. It may be defined as an assessment alternative to standardized, norm-referenced, multiple choice testing requiring active construction of meaning. Portfolios, performance-based assessments, exhibitions, journals, demonstrations, reflections refer to alternative assessment techniques (McMillan & Workman, 1998).

The new paradigm in teaching and assessment is communication-oriented, learner-centred, and process-oriented. It has generated open-ended, multiple solutions by integrating language skills and targeting to administer tests that also teach (Richards & Renandya, 2002). This changing perspective has led to an alternative perspective distinct from the traditional ones. Farhady (2003) regarded this change as a reform in education with emphasis on criterion related measurement. Thus rather than the traditional type tests, process oriented, task based, learner-centred, and strategy driven teaching and testing have been preferred. The paradigm shift, called assessment paradigm, assumes a greater role in teaching and learning process.

Educators have become more interested in a comprehensive qualitative account of the learners' ability via multiple measures in multiple occasions rather than a single score to make decisions (Farhady, 2003). In accordance with the new trends, the communicative principles are increasingly reflected in language tests (Taylor, 2006).

Alternative assessment is an umbrella term covering any alternatives to standardized tests (Assessment, Articulation, and Accountability, 1999, p.11). Therefore, it is not easy to define or categorize it. According to Garc'ia & Pearson (1994) performance assessment, authentic assessment, portfolio assessment, informal assessment, situated (or contextualized) assessment can be named as types of alternative assessment. According to Farhady (2003) alternatives to standardized tests can be categorized as alternative assessment, informal assessment, authentic assessment, performance assessment, descriptive assessment, and direct assessment. Performance-based assessments (projects, exhibitions, role playing, experiments and demonstrations), open-ended questions, writing samples, interviews, journals, learning logs, story of text retelling, cloze tests, portfolios, self and peer assessments, teacher observations and checklists can be listed as possible procedures of alternative assessment (Assessment, Articulation, and Accountability, 1999). The purpose of these types is not to find the correct response but to reveal learners' critical-thinking and evaluation skills through open-ended tasks by providing them with the opportunity to express themselves creatively (Zaimođlu, 2013).

### **2.5.1. Pros of Alternative Assessment**

From a positive perspective, Farhady (2003) summarizes certain characteristics of alternative assessment. First, it aims for the actual language use with authentic communicative function, thus much of alternative assessment is classroom based. It activates learners in their learning process. Second, it adopts a holistic

language view and stresses various aspects of language. Third, it reflects an integrative view of learners' skills and abilities. Fourth, it regards development of learning as a cognitive, social, and academic phenomenon. Fifth, it often collects information about the learner from a great variety of sources and by using different ways. Bayram states (2015) that alternative assessment highlights not only the process but also the product. He agrees with Farhady on its nature to reflect and simulate real life situations and claims that through alternative assessment practices, learners can have the opportunity to use creative and critical thinking skills. 'Alternative assessment' is usually taken to mean assessment procedures which are less formal than traditional testing gathered over a period of time rather than being taken at one point in time. They are usually formative rather than summative in function and often low-stakes in terms of consequences. These practices are also claimed to have beneficial washback effects. While Alderson & Lancaster (2001) mention the advantages of alternative assessment, they suggest that these provide easily understood information. They are also more easily integrated into the classroom despite some other weaknesses. Göçtü (2013), who has a positive opinion on alternative assessment, views it as a direct way of measurement for student performance and a trigger pushing learners to produce rather than reproduce what they know and generate rather than identify responses.

There has been an increasing positive approach towards alternative assessment and at the same time an increasing interest in research on this trend. The recent study results collected from various contexts have proved this tendency. Butler & Lee (2010) found out some positive effects of self-assessment on the students' English performance and also their confidence in learning English in their quantitative analyses. Bachelor (2015), who investigated high school student perceptions on



alternative assessment, found out that the experimental group students perceived the alternative assessments types favourably among three assessment types. The findings of Kavaliauskienė & Anusienė's research (2007) demonstrated that alternative assessment is a useful way for learner linguistic development. In a study conducted by Yüce (2015), the descriptive results displayed that pre-service English language teachers favour alternative assessment practices. In series of case studies carried out by Çetin (2011) teachers expressed very positive opinions about the impact of alternative assessment on learner autonomy, the affective and cognitive development of the students and their active role in learning process. These practices are claimed to promote differentiated learning and also motivate both learners and teachers. In another study on alternative assessment, Kızılkaya (2014) revealed that students mostly advocated the alternative types of assessment and found alternative assessment effective in terms of learner motivation despite declaring some concerns about the issue. The outcomes of Büyükduman's study (n.d.) indicated that students regard alternative assessment tools as useful. The research findings by Burnaz (2011) demonstrated that preparatory school intermediate level students had positive opinions about portfolios as an alternative assessment tool.

### **2.5.1. Cons of Alternative Assessment**

In spite of these promising research results of alternative assessment, there have been some concerns and criticism about these practices. In Kızılkaya's study (2014) students seemed to need more experience and time to have stronger opinions about alternative assessment. This demonstrates that they are not accustomed to these practices from their prior experiences. Teachers also need more training on alternative assessment procedures. Sometimes these procedures may not be appropriate or applicable in certain learning settings. For instance, in accordance with their study

results, class size is a vital factor affecting teachers' assessment preferences (Han, 2014; Özeren, 2013), which is a problematic issue particularly in Turkey. This factor may discourage teachers to implement alternative assessment. Due to the large number of students, it is not possible for the preparatory school students to take some roles in the assessment process (Taş, 2016). In Gökçen's study (2005) even though the majority of the instructors were satisfied with the use of projects as an alternative assessment tool, there were also some uncertainties and disagreements on the fairness of the project-based assessment. Most of the participants found projects to be less fair than traditional tests. Another major problem was about the implementation in terms of timing and assessing. Zaimoğlu (2013) states that objective techniques in classroom assessment are preferred by the teachers. From teachers' perspective, alternative assessment, which requires great effort in terms of planning and application, tends to be considered as a time-consuming assessment type and it raises some assessment-related issues such as objectivity in testing. Alderson & Lancaster (2001) stress its time-consuming procedures and regard these procedures as difficult to administer and score. The application phase of alternative assessment procedures in the classroom can be criticized since it is too demanding for teachers (Farhady, 2003). A more serious issue in terms validity and reliability of alternative assessment methods is considered among the current concerns in assessment (Clapham, 2000). Streff (2016) based on his research findings claims that the research of alternative assessments is inconclusive because of poor research design and lack of data even though some educators and researchers regard alternative assessments as a viable and a preferred method of assessing student learning.

In spite of all these disadvantages highlighted in the literature mentioned above, the effectiveness of alternative assessment on linguistic development, student

autonomy, creative thinking, and student motivation is supported by many study findings. Therefore, many practitioners consider these types of practices worth-trying to accomplish language learning. Even the most problematic testing factors of validity and reliability can be eliminated by the advantages of alternative assessment. For instance, Ana Huerta-Macías (1995) claims that alternative assessment includes valid and reliable procedures which avoid many problems derived from traditional testing such as norming and linguistic/cultural biases (as cited in Richards & Renandya, 2002). He adds that in the current literature, there is confusion about whether alternative assessment is an assessment, a learning method, or an artifact.

## **2.6. The Combination of Traditional and Alternative Assessment**

Considering the literature on traditional and alternative assessment, it can be stated that both assessment methods have advantages and disadvantages, which can differ from context to context. Although alternative assessment is an alternative or a modified version of its conventional rival as claimed by Streff (2016), in real learning-teaching contexts, they are commonly used as a combination rather than a sole separate assessment procedure. For instance, in Gökçen's study conducted in 2015, the instructors and administrators generally thought that some combination of projects and traditional testing should be used despite the fact that the majority of them favour project-based assessment, a complementary tool for the missing sides of traditional tests. Sönmez (2013) stated that practitioners are aware of the significance of the assessment and the ways of facilitating it in educational settings so effective teachers endeavour to make use of a variety of assessment methods and ways to determine students' needs and to achieve goals. Göçtü (2013) also highlighted his belief in the necessity of combination of traditional assessment with portfolio as an alternative assessment tool in order to emerge the advantages of both. In Çetin's study

(2011), the school assessment policy and the English curriculum support the use of multiple tools for on-going assessment. According to Cirit's research results (2015) on perceptions toward traditional, online and alternative assessment confirmed that the pre-service teachers believed not only in alternative methods but also TA practices to be used in combination. Supporting alternative assessment tools does not mean that teachers are expected to use them exclusively and it also does not mean that there is no place for traditional assessment in the language classroom (Toumi Lafi, n.d.). Nasab (2015) stresses the variety in assessment. According to him, teachers must attempt to put the diverse ways of assessment into practice and not to count on a single method of assessment too heavily.

Each learning context has its own characteristics. Even the assessment preferences of teachers and learners do not match with each other and also there are a number of factors influencing their decisions. The studies related to students' assessment preferences reveal that assessment formats which reduce stress and anxiety are preferred mostly by students (Doğan, & Kutlu, 2010; Büyükkarcı, 2010). However, there are some other factors such as gender and nationality according to study findings presented by Taş in 2016. The preliminary results of the study administered by Dittmann-Domenichini, Halbherr & Schlienger (2014) show that students are surprisingly honest and fair in assessing examinations independent of their own performance in the exam. From teachers' perspective, teachers facilitate assessments to help students enhance their learning and higher order skills and they prefer using objective techniques in classroom assessment (Zaimoğlu, 2013). However, there can be some other factors affecting their assessment choices. In a study conducted by Han (2014) investigating Turkish EFL teachers' assessment preferences and practices, it is found out that while class size is a significant factor on

these preferences, getting training on the relevant assessment method does not influence their choices. As a result, the assessment preferences vary by students and teachers.

The learning context itself is another factor on determining the appropriate assessment method. Hamp-Lyons (2007) defined two conflicting cultures of assessment: a learning culture and an exam culture. A learning culture aims individual learners' progress in learning whereas an exam culture targets learners' mastery of language proficiency in relation to that of groups or norms. The transition from an exam culture to a learning culture is a complex process, and teachers' views should be taken into consideration for a successful transition. According to this definition, assessment culture in Turkey, where this study is carried out, can be classified as exam culture. The sudden shift from one style of assessment to another might harm more than help. In order to achieve such a change, advanced countries have worked for over a decade to educate teachers, learners, parents, and community (Farhady, 2003). A great number of recent studies on performance-based assessment initiated by Ministry of National Education in Turkey may indicate the shift from TA to more contemporary assessment methods. However, the ongoing practices reveal that these methods are still regarded only as a complementary tool to already existing TA practices. TA practices have not yet been totally replaced with alternative ones in this context.

## **2.7. Performance-based Assessment**

As discussed above, there has been a paradigm change in language teaching pedagogy and practices towards a more communicative approach and correspondingly this shift has led to alternative language assessment procedures instead of the traditional ones. It is also suggested by Canagarajah (2006) as a change of emphasis;

in a broader perspective, from language as a system to language as social practice, from grammar to pragmatics and from competence to performance. Due to the emergence of CLT approach, alternative forms of assessment, including performance-based assessment have appeared (Kadidja, 2015). Recently in language teaching institutions PBA has been incorporated to enhance the language program. Putting the administration of such assessment forms under the investigation of program evaluation is quite significant in order to shed light to the other institutions in similar contexts.

Performance-based assessment (PBA) or performance assessment is not easy to be defined as a term or to be classified as a type of assessment. The more complete version of 'performance assessment' is actually 'performance-and-product assessment' (Fitzpatrick & Morrison, 1971) since performance assessment may contain both performance and product of students. Considering the nature and structure of performance tasks administered in this study, 'performance and product assessment' term is more suitable. Performance-based assessment is identified as an independent assessment type by some authorities while it is defined under the umbrella term of alternative assessment (Assessment, Articulation, and Accountability, 1999) by some others. From a broader perspective, alternative assessment may represent any and all assessments which are different from standardized tests (Marzano, Pickering & McTighe, 1993). Despite the fact that these are distinct terms with different meanings, alternative assessment, performance assessment, and authentic assessment are sometimes used synonymously. The term authentic assessment, introduced by Grant Wiggins in 1989, gives students the opportunity to apply their language knowledge and skills in the same way they are used in the real world, out of class. Performance assessment can be considered as a

broader concept than alternative assessment, including most of the characteristics of both authentic assessment and alternative assessment (Mitchell, 1992; Brooks, 1999). However, the most commonly confusing usage is that some describe performance assessment directly as alternative assessment regardless of any distinction. PBA is described as an alternative assessment, which requires learners to form a response, show implementation of knowledge, or create a product in authentic context (Authentic Assessment for English Language Learners, p. 239). Kavaliauskienė & Anusienė (2007) count PBA under the title of alternative assessment by stating that it can include oral presentations, portfolios, essays, and demonstrations. According to Büyükduman (n.d.), alternative assessment may include but might not be limited to authentic assessment, portfolio assessment and performance assessment, based either on a project or a problem in which learners reflect their knowledge via their performance. McMillan & Workman (1998) also classified PBA as one of the alternative assessment techniques including exhibitions, journals, demonstrations, reflections. Owing to the nature of performance assessment focusing on performance and authenticity, it may be confused with alternative or authentic assessment. From various different perspectives, PBA can be broader or more limited term compared to alternative assessment. Despite all these conceptual ambiguity, PBA has its own place theoretically in the literature and also practically in the language classroom.

In general performance-based education focuses on not only the process learners experience while dealing with a task but also the end product, encouraging them to make decisions and solve problems during their learning process (“Performance-Based Teaching and Assessment”, n. d.). Therefore, performance assessment’s main aim is to assess and monitor a student’s performance through a task or a project, which can be observed and assessed as directly as possible (Kadidja,

2015). PBA guides students' classifying, analysing and evaluating skills and trains them to have a control over their intrapersonal skills, time management and scheduling (Bayram, 2015). PBA is applied as various different types in accordance with different demands of language teaching contexts. Rather than offering paper-and-pencil tests, PBA typically consists of written production, oral production, open-ended responses, group performance, integrated performance (across skill areas), and other interactive tasks (Brown, 2004). In a performance assessment learners demonstrate their knowledge and skill by putting activity into practice or creating a product as lifelike as possible (Göçtü, 2013). Such assessment includes projects, interviews, presentations, critical thinking responses, or demonstrations (Büyükduman, n.d.). According to Grabin (2007), performance-based tasks, portfolios, journals, diaries, projects are some forms of PBA. Chen (2007) claims that PBA forms may range from the fairly traditional ones such as interview and essay writing to more recent ones such as communicative pair-work tasks, problem-solving tasks, group discussions and role playing.

### **2.7.1. Performance-based Assessment Tasks**

Nunan (1989) stated that a task focuses on meaning, rather than grammar and form by giving students the opportunity to interact in the target language (as cited in Bachelor, 2015). Practitioners adopting the communicative language method would be in favour of Nunan's view of task-based learning and assessment which became popular by the late 90's (Bachelor, 2015). Performance assessments are formed by sets of tasks putting emphasis on learners' performance skills rather than their proficiency skills (Gökçen, 2005). PBA includes two parts; a task and a list of scoring criteria or a rubric. The performance assessment task (PT) can be a performance, product, or extended written response, requiring critical thinking skills. Through these



tasks critical thinking, communication, and problem-solving skills, which cannot be measured via traditional assessment methods, can be assessed (Perlman, 2003). PTs enable students to show their ability to use and integrate knowledge, skills and work habits into a meaningful activity by reflecting their practical language skills in a real-life situation rather than their theoretical knowledge (“Performance-Based Teaching and Assessment”, n. d.). Students are active participants of their learning process by integrating, synthesizing, comparing and practicing information while the teacher acts as a facilitator. Assessment of PTs needs to be in multiple modes and measures and the evaluation should not only focus on language but also on content (Spector-Cohen, 2007). PTs, which can be designed to provide assessment information for any and all content and lifelong learning standards, are the backbone of a performance assessment (Marzano, Pickering & McTighe, 1993). The advance organizers in performance assessments are life-skill objectives and content-related performance tasks, plus ways that their achievement can be demonstrated in practice. While assessing performance tasks, performance of individual students and groups of students are mainly compared to a model performance. Set rubrics are used for grading and the sources of questions are typically selected from life-skill tasks and content specifications in curricular materials. In terms of program evaluation, there are some essential stages to construct a testing environment based on PTs; (1) identify areas of skills, (2) choose the type of assessment device, (3) design the assessment tasks, (4) determine scoring rubrics, (5) define standards for measuring performance, (6) train and arrange scorers, (7) validate the measures, and (8) administer, score, interpret, and report the results (Stufflebeam 2001).

PBA consists of two components: the performance task and a scoring rubric (Perlman, 2013). A scoring rubric can be in two forms: Analytical and Holistic

Rating. A rubric with two or more separate scales divided into sections related to hypothesis, procedures, results, and conclusion is called an analytical rubric. A scoring rubric using only a single scale is called global or holistic rating. Holistic scoring is often more efficient, but analytical scoring systems generally provide more detailed information that may be more useful in planning and improving instruction and communicating with students. Whether you choose an analytical or holistic rubric, you must clearly label and define each point on the scale. There is no best number of scale point, while holistic scoring is often simpler and faster than the analytic one. If assessment's aim is to provide information to guide program improvement, a quick overview of achievement might be particularly suitable for program evaluation to detect students who need more help and to assign final evaluations. However, both diagnostic value and efficiency can be optimized through concurrent use of analytic and holistic strategies (Herman, 1992). Perlman (2013) claims that a rubric can be a powerful communication tool. When shared among stakeholders, the rubric informs everyone about what characteristics of student work are most highly valued. It provides a means to clarify the vision of excellence and to convey that vision to the students. It can also provide criteria for assigning scores to subjectively scored assessments. She suggests sharing the rubric with students is essential if they are expected to do their best possible work. An additional benefit of sharing the rubric is that students adopt a critical eye on their own work. In order for a rubric to be effective, it must be understandable for the interlocutors. This may require restating the rubric, eliminating metalanguage and explaining the criteria in a way that is proper for the learners' developmental level. Practitioners interested in using rubrics to assess performance-based tasks have some options; use an existing rubric as it is, adapt or combine rubrics to suit a specific purpose, or create a rubric

from scratch (Perlman, 2013). Rubrics used primarily in alternative assessments have replaced the traditional answer keys (Streff, 2016). By using set rubrics, grades are assigned to each student's performance and it enables assessment of the quality of their achievements (Stufflebeam, 2001). Making comparisons can be eliminated through the use of rubrics to evaluate the tasks. Scoring rubrics have an important impact on PBA, for instance, a well-formed analytical rubric can minimise some of the weaknesses of authentic assessment related to intra and inter-rater reliability (Kadidja, 2015). Rubrics are designed for analytic scoring in order to provide teachers with rich, detailed information and through these kinds of rubric student performance on specific task elements are measured (Moon, Callahan, Brighton & Tomlinson, 2002). Cohen (1995) states that developing good evaluation criteria is a complicated and challenging process but also provides assessing the right knowledge or skill. However, the findings from the questionnaires show that only half of the teachers use rubrics although the Ministry of National Education in Turkey (MONE) encourages teachers to utilise them (Yıldırım & Orsdemir, 2010). Unlike a traditional grade, which summarizes all aspects of pupils' performance in a single number, letter or word, a rubric provides information on pupils' performance on each of the criteria. This gives a more general profile of pupils' ability for formative and summative purposes ("Performance-Based Teaching and Assessment", n. d.).

### **2.7.2. Pros of Performance-based Assessment**

Considering the positive aspects of PBA, it has been advocated by many experts and practitioners for some reasons. PBA activates and facilitates higher-level thinking and problem-solving skills (Spector-Cohen, 2007). During the evaluation and feedback phase of PBA, an interaction and effective communication arise between the teacher and the student enabling positive impact on the teaching and learning process.

However, teachers usually regard the students' exam scores more than their performance (Zaimoğlu, 2013). Stipek (2002) revealed that assessing student performance increases the self-confidence of the student. Özeren (2013) states that PTs focus on not only knowledge but also the real performance of the learner through problem-solving skills. He also stresses their feature of being process-oriented. It is one of the distinct characteristics of PBA, different from the more traditional assessment methods. According to Brown, despite demanding extra efforts, the form of more direct assessment including actual performance or simulated real-world tasks compensates its downsides. Higher content validity is achieved due to the assessment in the process of performing the targeted linguistic acts. Through interactive tasks communicative performance of the student is assessed, which is not elicited by paper-and-pencil tests (Brown, 2004). PBA can create rich learning contexts, trigger students' critical self-evaluation, simulate real-world problem solving, raise practitioners' awareness of their students' cognitive processes and lead to good instruction through excellent assessment of students' abilities to solve, evaluate, and synthesize problems. For practitioners, learning how to use the scoring rubric of PBA can be an effective staff development experience. Some learning outcomes cannot be assessed by other assessment formats but PBA (Perlman, 2003). Performance-based assessment has been regarded as a promising assessment which reflects students' communicative competencies truthfully among other formative assessment methods (Chen, 2007). Brown & Abeywickrama (2010) consider PBA the same as alternative assessment and claim that it enables students to use the target language for communicative purposes via open-ended and authentic activities. These activities assessed through PBA have high authenticity, washback and validity. Brooks (1999) predicts that PBA will continue to play an increasingly important role in many

English language programs since it reflects curricular objectives, and has positive washback effects. These assessments result in an observable performance or concrete product by encouraging revision and self-evaluation. Scoring PBA requires judgment and scoring criteria is shared by the target audience (Marzano, Pickering & McTighe, 1993). Moss (1992) claims that compared to the consequences of traditional assessment alone, implications of performance assessment are more beneficial to learning and teaching and provide more valid interpretations of certain complex educational domains.

### **2.7.3. Cons of Performance-based Assessment**

In spite of its positive aspects, performance-based assessment and its tasks in practice are criticised by some others. For instance Önal (2010) states that in large-scale testing performance-based elements are often limited to a small number of controlled task types involving speaking and writing since they are costly and they demand significant resources to administer. Brown (2004) considers PBA as time-consuming and therefore expensive. Providing the necessary conditions is usually difficult to employ the performance testing approach and also requires a great deal of time and resources for development and implementation (Stufflebeam, 2001). Perlman (2003) agrees with Stufflebeam that these kinds of assessment are expensive and time-consuming and adds that considering the assessment results, a good result on one PT may not be generalized well to similar tasks. The subjectivity in scoring may make some practitioners uncomfortable, which can be hindered with a well-constructed rubric, effective rater training. According to Perlman (2003), other assessment formats, such as multiple-choice tests can measure certain kinds of knowledge and skills more efficiently than PBA. Messick (1992) questioned the positive washback and validity aspects of PBA, which are mostly regarded as the

positive features of this assessment form. He believes that there are many reasons why direct performance and authentic assessments do not really produce positive washback. In terms of construct validity, the concern is that PBA should provide representative coverage of the processes and content of the construct domain in order to guarantee the score interpretation not to be limited to the sample of assessed tasks but to be generalizable more broadly. Moss (1992) also stresses the validity problems of PBA, which are not easily handled with traditional approaches and criteria. He also urges that these forms of assessment present fewer independent and complex responses although they create extreme freedom for students in interpreting, answering, and even sometimes designing tasks. Furthermore, the evaluation phase of this kind of assessment requires expert judgment. In their paper Linn, Baker & Dunbar (1991) have analysed some technical concerns in testing aspect of PTs such as fairness and validity and also questioned the scoring criteria, task numbers, their selection and complexity in detail. They believe that for teachers providing training and support on PBA is essential. Metin investigating teacher preparation of performance assessments revealed that teachers had issues in preparing and administrating PBA (Metin, 2013 as cited in Streff, 2016).

#### **2.7.4. Research on PBA**

Considering the studies on PBA or particularly the studies investigating the effect of PT implementation, there have been both positive and negative results in terms of the stakeholders' views. For instance, in a study, which focuses on the view of administrators and teachers on project-work usually categorized under the title of PBA, conducted by Gökçen in 2015, the majority of the teachers seem to be in favour of using project-based assessment and generally view it as a complementary tool for the traditional tests. The administrators' views revealed that due to the limitation of

traditional tests for measuring students' language proficiency, they encouraged the use of projects in their curriculum. In a study conducted with Taiwanese college students by Chen (2007), students generally showed positive responses to tiered PTs presented in a final examination for English listening and speaking class. The positive sides observed were appreciation of choices of tasks, improved English skills, heightened motivation, greater confidence, and increased effort. Negative aspects of PTs were partnership time required to complete the task, scoring and complexity level. Positive findings reveal that in promoting English language learning of college EFL learners in Taiwan's differentiated instruction is promising. Kadidja (2015), from Minnesota State University, carried out an MA dissertation with ESL university students examining the effect of performance-based assessment on their motivation. According to the research findings, the students responded positively to this type of project even though their emotional and motivational states altered across time related to their experience with their performance, the cohesion of their group and the oral presentation. On the other hand, the results of a longitudinal study (Chinda, 2014) investigating the reactions of tertiary level Thai EFL teachers towards the use of PBA reveal that teachers adopt quite negative attitudes toward PBA after a six-year implementation. The weaknesses they mentioned are particularly related with the tasks, rater training and rating scales.

In Turkey, there have been a great number of studies conducted on this issue with young learners (YL), young-learner teachers and parents due to the reforms in the English Language Curriculum for Primary Education. In this new curriculum, dated February 10, 2006 with the resolution number 14 and put into practice by Board of Education, the use of both summative and formative assessment tools were encouraged. As a formative assessment method, the use of performance tasks was one

of the innovations put into use by this new curriculum developed by the Ministry of Education in Turkey (Ak, 2013). After this innovation, researchers focused on the impact of PT implementation mostly in terms of stakeholders, often disregarding the administration component. Due to the large scale usage of PTs, the implementation was put into practice for each type of course in primary and secondary education and studies were predominantly carried out in maths, social sciences, Turkish language courses and English language courses. However, these studies, which are predominantly focusing on young learners, are limited in terms of the age. As an example, in Ak's study (2013), according to the data collected from YL teachers, students and parents, the PT implementation is generally viewed as a useful method, which improves students' language skills and research skills. These stakeholders also emphasized the fun factor of this implementation. However, the number of tasks, the task complexity and lack of knowledge were the downsides of the new implementation. In another study carried out with 43 young-learner teachers from ten schools by Yıldırım & Orsdemir (2013), the results display that despite vocabulary improvement, grammar reinforcement and some affective aspects such as increased motivation and improved confidence, PTs have weaknesses in certain areas according to the respondent teachers.

Gönül (2010) conducted a study in a private college with a small number of young learners, YL teachers and parents. Dependent on a specific course, she focused on the general effect of PTs from the participants' point of view and the results indicated that all participants agreed on the contribution of PTs on student success. While parents demanded more detailed feedback, students wanted to be involved in the PT preparation process themselves. Although the study findings revealed that PT implementation improves higher-order skills, it was also seen that it takes quite a long



time. From a different aspect, Paksoy (2015) focused on the compliance of ethical rules of middle-school students while doing their performance tasks in her study. The researcher detected some concerns about the autonomy aspect in PTs due to the students' tendency for copying and pasting information from the internet and their lack of knowledge about intellectual property. PT contribution is investigated not only in English courses but also in the other courses such as social science and maths. For instance, in an experimental study conducted by Pulat (2014) with secondary school students, the findings indicate that the students who have experienced PT implementation showed a significantly positive difference in terms of success and critical thinking skills compared to the control group. PT practice in maths lesson is also investigated by Hacısalıhoğlu (2013) in a research, focusing on the student opinions and their challenges in the PT process, executed to young learners. The study results display that students are more aware of the application phase than the learning outcomes of PTs. The downsides of the PT practice are observed as the inefficiency of group-work, unawareness of resource use and tendency in overuse of the internet by students.

After examining the downsides and pluses of PBA thoroughly, within the frame of the current context one should plan well how to integrate this assessment method into the existing language program and provide its consistency with that program or may ponder whether to integrate it or not since the stakeholders and their attitudes toward performance-based assessment play a significant role on the success of PT practices. The present study aims to evaluate the PT implementation process in accordance with the current language program carried out in the target institution.

## CHAPTER 3 METHODOLOGY

In this chapter the context, participants and the language program of the preparatory school under the investigation of this research are presented in detail particularly by describing the assessment procedures before, during the trial year, and after the PT implementation as the main focus point of the study. The data collection instruments and procedures as well as data analysis procedures are exhibited.

### 3.1. Context

This study was carried out in fall 2017 semester at a preparatory school of a state university in Marmara Region of Turkey, which represents an EFL context. The skills-based education program of the preparatory school has been implemented for five years, in which the skills and language elements are instructed as separate courses and assessed through discrete-point tests. Although only listening and speaking skills are taught under the heading of one course, its assessment practice is still discrete-point. There are five courses, namely listening/speaking, reading, writing, grammar and vocabulary. The curriculum is instructed through course books and supplementary materials designed by instructors to meet the needs effectively. The skills-based education program is first introduced to this prep-school by an academician from English Teaching and Education Department, who was also the director of the preparatory school at that time. The program has been redesigned, adapted and updated in time under the supervision of the administration and with the help of teacher development and curriculum design committee in accordance with the new trends and developments in ELT. In order to trace the new paradigm in assessment, PBA is considered as a supplementary tool to TA practices. By the time of the first trial of PTs, assessment has been carried out predominantly through TA

methods. As all other new implications in the context, the performance assessment tasks (PTs) are designed in the light of mutual agreement reached at regularly scheduled contact meetings in terms of observed learner needs.

The objective of this preparatory-school education is to get students to attain B1 proficiency level of English in all language skills, to provide them with the essential oral and written skills and strategies to utilize in social life, to get them to master four language skills in order to be competent in the target language in various fields, to teach the students not only the basic grammatical rules of foreign languages but also the comprehension and production in these languages according to the Common European Framework Certificate and to teach the language for the learners' academic goals in their future education and business life. One-year preparatory education is carried out as compulsory or voluntary.

The program is composed of three basic language levels. These levels are determined through the placement test at the beginning of the academic year. Course materials vary according to language levels. However, there are common course books for three levels. Although the exit level is officially determined as B1 level, pre-intermediate and intermediate levels attain B2 level at the end of the year. Discrete-point achievement tests for each level are prepared in accordance with the students' proficiency level. At the end of the year, all levels are administered the same test in B1 level. Modular system, in which students are supposed to take a test to pass each module, is not carried out.

The weekly course hours differ in terms of language levels. Students at elementary level are taught 26 hours of English per week. While students at pre-intermediate level are taught 24 hours of English per week, students at intermediate level have 22 hours of English courses per week.

Weekly course hour distribution varies by language levels. At elementary level listening and speaking course involves 7 hours, grammar involves 7 hours, reading course involves 5 hours, writing course involves 5 hours, and vocabulary course involves 2 hours. At pre-intermediate level listening and speaking course involves 7 hours, grammar course involves 5 hours, reading course involves 5 hours, writing course involves 5 hours, and vocabulary course involves 2 hours. At intermediate level listening and speaking course involves 7 hours, grammar course involves 4 hours, reading course involves 5 hours, writing course involves 4 hours, and vocabulary course involves 2 hours.

Specific learning outcomes of each course are determined by the administration, improved by the relevant coordinators and put into practice by the practitioners. Both the course outcomes and the general instructional goals of the language program are monitored, revised and updated in harmony with the curriculum learning materials and student needs. The written forms of these procedures are prepared for the upcoming accreditation process of this institution and available in the preparatory school's web-site. The decisions are made through contact meetings and recently through qualitative, quantitative or mixed research techniques in respect of stakeholders' opinions and in the light of new developments in ELT. Any component of the language program such as assessment, learning materials or schedules might be under investigation. The instructional program is evaluated in the target institution every three years.

### **3.2. Participants**

In total 129 EFL students, 60 instructors, and two administrators participated in this study.

### **3.2.1. Students**

In the study, while 126 students participated in the initial open-ended questionnaire, the scale was administered to 129 students. Participant students in the study were randomly selected EFL students from three different language proficiency levels, whose ages ranged from 18 to 30 years. The classes consisted of both genders. Only the students from daytime education participated in this study. Participants received formal English education at state and private high schools for different durations. At the beginning of fall 2017 semester, they all took a placement test tailored by the testing unit of the institution in order to be placed in the suitable level of English language proficiency. The participant students were from various departments including engineering, psychology and chemistry. In some of those, one year English prep-school was compulsory while voluntary in others.

### **3.2.2. Instructors**

The number of the participant instructors was 60. Participant instructors in the study were serving at various language levels such as elementary, pre-intermediate and intermediate level at Preparatory School in EFL context. The participants completed their BAs at state or private universities in Turkey. Almost half of those had an MA degree. Three of them were still doing their PhDs during the study. Three participants were foreign instructors, two of whom were native speakers. Except for one instructor, all instructors had more than five year teaching experience, mostly in the same context. The participant EFL instructors taught English approximately twenty hours each week to adult learners. All of them had a specific duty at least in one committee such as testing unit. Most instructors had experience on testing. The whole staff participated in this research.

**Table 1.** The EFL Instructor and student distribution by language levels

Daytime Education		
Level of English	Number of instructors	Number of students
Elementary	40	929
Pre-intermediate	21	243
	(inter-levels alternately)	
Intermediate	21	82
	(inter-levels alternately)	
<b>TOTAL</b>	<b>61</b>	<b>1254</b>

### **3.2.3. Administrators**

Two members of the executive staff, namely the director and the assistant director, were also included in the research as stakeholders. When an evaluator invites as many stakeholders as possible, the diversity of interest will broaden (Weiss, 1998, pp. 103). The director was a professor at the Faculty of Veterinary while the assistant director was an instructor from the preparatory school German Department. The researcher, who was one of the assistant directors of the institution, was not one of the respondents for the sake of the objectivity of this study.

### **3.3. General Assessment Procedures in the Context**

Assessment is carried out by the testing unit consisting of nine members including the head of the unit. All tests are checked by the relevant coordinators and the assistant director before the delivery. The discrete-point tests, mid-terms (6) and quizzes (6), are designed to assess students' different language skills and elements separately. These are mostly paper-based, in both formative and summative forms. In the proficiency exam students are not tested orally. To take the proficiency exam at the end of the academic year, students' GPA must be minimum 60. To be able to pass the proficiency exam, exam scores are calculated as follows: 50% of annual average + 50% proficiency exam and the average must be minimum 70.

### **3.3.1. Assessment Procedures before the PT implementation (Fall 2016 Semester)**

All the procedures explained in the general assessment procedures mentioned above were valid in the fall 2016 semester. Beside the TA practices, there were three performance-oriented and process-oriented applications, namely a Speaking Video Project related to the listening/speaking course, Extensive Reading application related to the reading course and writing evaluations related to the writing course. Except for the writing evaluation, administered twice a term, the other two were not content-based assessments. Writing evaluation was administered as one shot TA by giving students one topic to write on in limited class time. In extensive reading application, students were reading five books per term and assessed through two traditional multiple-choice tests including 2 or 3 books. In speaking video project students were assigned a topic and as a group-work they shot a video about that topic. In total these three applications composed the 20% performance score of the general score average as presented in Table 2.

The number and intensity of traditional exams, failing to measure the actual performance of students and their failure to transmit students' structural and lexical knowledge into written and oral production have been criticized by the instructors in contact meetings. It has also been discussed that students made no effort to practice English out of class and did not develop learner autonomy. Except for the summative speaking project application once in a term, writing evaluation and extensive reading practices, aimed to measure performance and expected to be more process-oriented, failed to measure real performance since they were assessed through traditional exams rather than performance-based assessment methods. In this respect, these were found insufficient by the practitioners. Alternative assessment practices to solve these problems have been sought to compensate for at least one pen and paper quiz.

**Table 2. The assessment percentages in Fall 2016 Semester before PTs**

Assessment Type	Frequency per semester	Total Number of Exams	Percentage
Mid-term	2	2 x 6=12	50 %
Quiz	2	2 x 6=12	30 %
Performance-oriented applications extensive reading test (6 points) + writing evaluation (8 points) + speaking project(6 points)	1	1 x 3= 3	20 %

### 3.3.2. PT Implementation Trial (Spring 2017 Semester)

As seen in Table 3, the traditional exams and performance-oriented applications and their scoring percentages remained the same in spring 2017 semester, when PT implementation was first introduced on volunteer basis to the students as a trial. The tasks and rubrics were developed from various sources based on literature readings or originally generated by the relevant coordinators. The preparation process took three months. The application and scoring took three weeks. The trial implementation was limited to the grammar and vocabulary courses. Only volunteering students participated in the PTs to get 5 bonus points as part of their last relevant mid-term exam.

It was observed that the number of volunteers was more than expected and the tasks prepared were promising. Therefore, these positive results encouraged the institution to continue the implementation on a larger-scale. In the annual evaluation meeting, the number of the paper-based exams was discussed and consensus was reached on compensating for one pen and paper quiz of four courses, namely listening/speaking, grammar, reading and vocabulary courses with a PT in the



following academic year. The writing evaluation practice was replaced with writing portfolio after the discussions, yet it is not in the scope of this research.

**Table 3. The assessment percentages in Spring 2017 Semester (PT Trial)**

Assessment Type	Frequency per semester	Total Number of Exams	Percentage
Mid-term	2	2 x 6=12	50 %
Quiz	2	2 x 6=12	30 %
Performance-oriented applications (as described for Fall 2016)	1	1 x 3= 3	20 %
Optional Performance Assessment Tasks (Grammar & Vocabulary)	1	1 x 2= 2	5 points bonus

### 3.3.3. Assessment Procedures during the PT implementation (Fall 2017 Semester)

All the procedures presented in the general assessment procedures were valid during the actual implementation of PTs. In addition, performance-based assessment tasks plus portfolio assessment formed the 20% of the general average score as seen in Table 4. According to this new implementation, different from the first trial, the number of PTs was increased and number of quizzes was decreased. They were applied in accordance with the content of not only grammar and vocabulary courses but also listening/speaking and reading courses. On a greater scale, all the students were included in the PT practices compulsorily. The tasks were scored on 4% for each course. To sum up this new PT implementation became a significant assessment component of the whole assessment system. Portfolio assessment as an individual assessment type was not investigated in this study as the focus point of the study was mainly performance assessment tasks. Moreover, as the present research intended to

be a formative evaluation, merely the fall term procedures were analyzed. However, the outcomes of the study were evaluated and put into practice for the sake of the ongoing language program improvement in Spring 2018 Semester.

**Table 4. The assessment percentages in Fall 2017- Spring 2018 (Current PT implementation)**

Assessment Type	Frequency per semester	Total Number of Exams	Percentage
Mid-term	2	2 x 6=12	50 %
Quiz	1	1 x 6=6	30 %
Obligatory Performance Assessment Tasks (Grammar /Reading/Vocabulary/ Listening & Speaking)	1	1 x 5= 5	20 %
16% + Portfolio Assessment (Writing) %4			

### 3.4. Current PT implementation in the Context

#### 3.4.1. The Objective

One of the main objectives of the implementation of performance assessment tasks in this institution was to lessen the ratio of traditional paper-based assessment whose preparation, administration and grading phases were a serious workload for the instructors. In addition, instructors observed that students could not transfer their structural or lexical knowledge into written and oral output. They were generally weak at productive skills. Through the performance-based assessment tasks, in accordance with the objective of the language program, it was aimed to enhance their productive skills plus the receptive ones. PBA, by their nature, were formed to foster all language skills in an integrated way. For instance, for a reading PT, students were expected to read a book, write a book report, present it in the classroom and also listen to the other students' presentations. Four language skills were activated and aimed to

improve by this way. It was also aimed to get students to improve their autonomy and higher-order skills through motivating tasks that were prepared out of the class. In EFL context, usually the only environment the students use the target language is the classroom. The current PT implementation was designed to give the students the opportunity to practice English out of the class as a part of their learning process.

### **3.4.2. Planning**

Based on the dense student participation of the first trial year and in accordance with the consensus reached in the evaluation meeting, coordinators were assigned to revise the tasks used previously and also to develop new tasks through some activity books such as Grammar Games and Activities (Wacyn-Jones & Howard-Williams, 2001), Vocabulary Games and Activities (Wacyn-Jones, 2001), and Instant Lessons (Howard-Williams, Tomalin, Wacyn-Jones & Woods, 2001) and some useful web-sites. The rubrics used were also improved and some of them were adapted from other samples. Throughout the planning process, coordinators did literature reading on performance-based assessment and they consulted with their lecturers in MA or PHD programs for the final versions of tasks and scoring rubrics. Coordinators worked on those with teams of instructors for almost two months during the summer holiday. Finally, the tasks were handed in to the administration for proof-reading. The trainings, seminars and contact meetings were arranged. The weekly schedule for the PTs was prepared. Some extra training sessions on alternative assessment were planned by the professional development unit of the institution.

### **3.4.3. Performance Assessment Task Descriptions**

Four new PTs were designed for the reading, grammar, listening/speaking and vocabulary courses for the 2017-2018 academic year. Except for the PTs designed for

the reading course, all others were designed parallel to the content of the courses. While PTs for reading course consisted of two tasks per semester, students prepared one task per semester for the other courses. Because they compensated for one paper-based quiz, the content was limited to the subject matter taught for three weeks. Each course had its own scoring rubric for its PTs. Vocabulary and reading PTs did not vary from level to level whereas there was a level distinction for the grammar and listening/speaking PTs. According to the syllabus, students were taught different grammatical structures depending on levels. Therefore, elementary level students prepared different tasks from pre-intermediate and intermediate levels since the tasks were created in contexts related to the relevant grammar structures. Different from the others, the PTs designed for listening/speaking course were mainly based on student interaction to be done as pair-work and group-work tasks. All the PTs were presented with a variety of options and students were free to select among them.

#### ***a. Reading Course PTs***

PTs administered in reading courses were not parallel to the course content since they were designed as extensive reading (ER) applications by adhering to extensive reading principles (Appendix D). The reading PTs were mainly based on ER activities generated by Bamford and Day (2003). Six performance tasks to be prepared were related to the two graded books the students read and these consisted of tasks such as writing a book report, writing a summary and making a story map, preparing a quiz show, role-playing a scene from the story, making character profiles, and comparing the book and its movie. The learning objectives of these tasks are reading for pleasure, thinking critically, summarizing, sequencing, role-playing, and comparing-contrasting. They all require individual work. The determined format of PT presentations was via Power Point presentations (PPP) or posters. The students

were free to select the books to read but they were presented a guide-chart by their reading instructors to choose the right language level (Appendix A). The instructors were given an instruction paper of ER PTs and an evaluation form.

#### ***b. Vocabulary Course PTs***

PTs administered in vocabulary courses were based on the three units of the book used in the course. There were 16 PTs and each term the students would choose one PT. These consisted of various tasks (Appendix D) such as preparing a quiz show, forming a word map, preparing a wheel of fortune. Students were supposed to select 15 words from the relevant units to prepare these tasks. All these PTs were designed as individual work. Some of the learning objectives of these PTs were reinforcing target vocabulary, using words in a wider context, specifying the target vocabulary, and making use of lexical skills through intercultural awareness. The process was to be shot as a video recording. The determined presentation format of PTs was Power-Point presentation (PPP), poster, wheel of fortune, or video shooting.

#### ***c. Grammar Course PTs***

PTs administered in grammar courses, consisted of three units of the grammar book, were content-based. Five tasks for elementary level, six tasks for pre-intermediate and intermediate level were generated. Students were supposed to incorporate five sentences for each target structure to their PTs to make up a total of 15. The tasks (Appendix D) were taken from various activity books, useful web-sites ([www.teachingenglish.org.uk](http://www.teachingenglish.org.uk)) and were adapted to the context. Some of the learning objectives of grammar PTs were using specific structures in a defined context, improving critical thinking strategies, organizing ideas by using specific grammatical structures. All grammar PTs were performed individually by the students and presented as voice recordings or video shootings.

#### ***d. Listening and Speaking Course PTs***

The PTs generated for listening/speaking courses (Appendix D) were also based on the content of the course. All tasks, adapted from the course material ‘Communicate 1 / 2’ (Pickering, 2012), were prepared in pairs in order to foster student communication and interaction. Two different tasks were designed for each language level. Learning objective for each task is to enhance speaking skills of students. However, some high order skills such as role playing, thinking critically, working collaboratively were also targeted. The format of presenting PTs was writing a tape-script of the dialogue and shooting a video (a short film), limited to 5 to 10 minutes.

#### **3.4.4. Scoring Rubrics**

Both holistic and analytic scoring rubrics appropriate to the language skill, language element and the nature of the produced tasks were used. In accordance with the nature of performance assessment, all rubrics consisted of a section as ‘task completion or task achievement’. Among these scoring rubrics, the only holistic one was prepared for the reading PT in accordance with the principles of extensive reading. Except for the holistic rubric designed for extensive reading PTs, which was scored over four, all others were scored over a hundred. As there were two tasks assigned for the reading course, the average was calculated and turned into a hundred. All scoring rubrics were based on readings in the literature and some of them were adapted from other rubrics. Irrelevant products led to reduction of 10 points from the total score. Rubrics were prepared by the coordinators and their team. Their proof-reading was made by the administration. All PT scoring rubrics were shared with the students so as to inform them about the requirements (Appendix E).

The scoring rubric used to assess extensive reading (Appendix E) administered in reading course was developed by following the guidelines in Weigle (2002). Since the main aim of extensive reading is to read for pleasure, a holistic rubric between 1-4 points was found suitable. The criteria included task completion, relevant content and artistic creativity and originality. With a holistic rubric it was aimed to generally describe what a task should look like. By using those scoring rubrics, instructors scored the two PTs and took its average and turned the score into a hundred to reach the final score.

The scoring rubric used to assess vocabulary PTs, developed by the vocabulary committee following the instructions by Roberts (2008), was an analytic one (Appendix E). It was calculated over a hundred. The requirements of the rubric were task completion, accurate use, presentation, and creativity. Each scale was calculated over 25 points. In order to generate the rubric, a number of sample scoring rubrics were investigated on the internet. It was formed by an adaptation of these by determining the content areas and the scoring bends.

The scoring rubric used to assess grammar PTs was an analytic one with the most grading bends among others (Appendix E). Although there were four main requirements to score, the criteria 'accurate use of required structure' was repeated three times in the rubric to assess each grammatical structure separately. The scoring rubrics related to performance assessment in the literature were designed mainly for language skills, even more for productive skills. Therefore, it was quite difficult to find a suitable grammar rubric in the related literature to assess the performance of students. Grammar was mostly tested through traditional assessment methods. However, through performance-based assessment tasks, the structures learnt can be transferred to oral or written output as a performance. By following the guidelines

introduced by Weigle (2002) and adapting rubrics for productive skills, an analytic rubric was developed. The criteria were accurate use of required structure, content, task achievement, and organization. It was designed to be scored over a hundred.

The scoring rubric generated to measure the PTs of listening/speaking course was adapted from “Common European Framework speaking evaluation rubric” (Appendix E). The instructors were very familiar with the rubric since it was already used to assess the oral performance of listening/speaking course in the preparatory school. The 100-point rubric was turned into a 60-point one in order to make room for the items to evaluate creativity. In addition to the main requirements such as sentence structure, vocabulary, fluency, pronunciation, and task achievement, four other bands related to creativity, participation (in pair-work), visual aids and presentation skills were incorporated to the original rubric.



### 3.4.5. PT Implementation

The whole implementation process took 15 weeks as shown in Table 5.

**Table 5. Weekly Assessment Schedule at Preparatory School**

DATE	WEEKS	EXAMS
18-22 September 2017	1	Orientation to students
25-29 September 2017	2	Alternative Assessment training for instructors
2-6 October 2017	3	General assessment PPP for instructors and students
9-13 October 2017	4	General Quiz / Extensive Reading PT training
16-20 October 2017	5	Deadline to choose Extensive Reading Task 1/2
23-27 October 2017	6	GR Quiz / LSP PT training
30 October-3 November 2017	7	Speaking Quiz / WR Portfolio Submission 1
6-10 November 2017	8	V Quiz / V PT training
13-17 November 2017	9	WR Quiz / ER Task 1 Deadline / GR PT training
20-24 November 2017	10	1st MID-TERM
27 November- 1 December 2017	11	WR Portfolio 2 / Deadline to choose GR & V & LSP Tasks
4-8 December 2017	12	
11-15 December 2017	13	Reading Quiz
18-22 December 2017	14	Listening Quiz /GR / LSP/ ER Task 2 /V Deadlines / WR Portfolio 3
25-29 December 2017	15	PT Evaluation of completed task presentations
2-5 January 2018 2017	16	2nd MID-TERM

\*GR: Grammar / WR: Writing / V: Vocabulary / LSP: Listening/Speaking

#### *Week 1*

On the first week of the preparatory school, instructors informed students about the general procedures such as courses, attendance, classroom rules, and exams. In this orientation session, PT implementation was also basically presented as a part of the general assessment procedures.

### ***Week 2***

A seminar on alternative assessment, which lasted an hour, was held by the professional development unit for instructors in order to introduce alternative assessment and to give some essential information about non-traditional assessment procedures.

### ***Week 3***

A training session, based on the main principles and procedures of performance assessment tasks and its place in this preparatory school's assessment system, was provided for the instructors lasting for 2 hours including an interactive phase at the end. The whole assessment system adopted in this preparatory school was presented in Turkish (Appendix F). However, the PT section was highlighted more since the instructors were already familiar with the other procedures. At the same week, the supervisor instructors allotted one class hour for this presentation to give training to the students.

### ***Week 4-Week 9: Training weeks for both the instructors & students***

Each PT practice was explained at the contact meetings organized by the relevant coordinator to the instructors and related documents such as informative forms, tasks and rubrics (Appendix B, D, and E) were delivered. The contact meetings for instructors were regularly run at 2 p.m. on Wednesdays, which had been a routine meeting time for more than 5 years in the preparatory school. At the same week instructors allotted one class hour to the training of their students. The weekly schedule was announced to the instructors and posted on the billboards of the classroom for students to follow the dates and deadlines.

#### ***Week 4***

After explaining the main objectives and PT application procedures to the stakeholders, the process of implementation started with the first PT, extensive reading task I. First a contact meeting was held for the instructors by the reading coordinator. All the instructions, objectives and task descriptions and scoring rubrics were presented to the instructors and their questions were answered. The same training was carried out for the students in one class hour of reading course in week 4 and the deadline of reading task I was announced as week 9. Two-week time was given to students to buy their books and to determine their tasks for both PT I and PT II among the options. The students were firmly warned to complete their reading tasks beforehand so as to gain sufficient time for the other PTs.

#### ***Week 5***

The students selected their book names and task numbers so reading instructors noted their choices in the tracking forms. Therefore, choice of reading PT I and II was completed. The students were expected to complete PT I in week 9 and hand in PT II in week 15. Students had approximately five weeks to complete and hand in reading PT I and another 5 weeks for PT II.

#### ***Week 6***

The PT designed for listening/speaking course was first presented to the instructors in a contact meeting and respectively one class hour of listening and speaking course was allotted to train the students. The tasks were explained and the scoring rubrics were shared. Each session had interactive sections for the students' questions.

### ***Week 7***

The students were expected to write their first writing evaluation for their portfolio in the class. There was no PT training in week 7. The students were reminded about the deadline of the reading PT I by their instructors and also guided for listening/speaking PT.

### ***Week 8***

The PT designed for vocabulary course was first presented to the instructors in a contact meeting and respectively one class hour of vocabulary course was allotted to train the students. Each session had interactive sections for the questions on their minds.

### ***Week 9***

The students handed in their first PT, ER task I. It was scored according to the scoring rubric and the scores were announced in one week. Some of the instructors preferred to give feedback (Appendix G) on those PTs in order to guide their students in their following attempts. At the same week the PT designed for grammar course was first presented to the instructors in a contact meeting and respectively one class hour of grammar course was allotted to train the students. Each session had interactive sections for the questions on their minds. Therefore, the training sessions for PTs were completed and students were expected to decide on their tasks for listening/speaking, grammar and vocabulary courses until the first day of week 11. In this case, students had five weeks to determine their tasks for listening/speaking PT, three weeks for vocabulary PT and two weeks for grammar PT. The official deadline to determine their PTs was announced as week 11; however, the students were expected to initiate their preparation process, if they wished, before informing their instructors.

### ***Week 10***

The students took their mid-term exams for the whole week. They were also expected to think about their performance tasks during this period. Since generally the Turkish students come from a traditional education background and are not very familiar with process-oriented assignments or testing methods such as performance assessment, in week 10 just one week before the actual implementation a warning notice was posted on the classroom doors (Appendix A), which stressed the importance of attending the classes between week 11 and week 14 and extending their PTs over the period of the time given.

### ***Week 11-Week 14: Students' preparation period for the PT presentations***

After informing their instructors on the first day of week 11, students were expected to start working on their PTs. Between week 11 and week 14, for four weeks the students were supposed to follow the content of the courses carefully to prepare the content-based tasks, which were vocabulary PT, listening/speaking PT and grammar PT. The students were warned by their reading instructors to complete reading PT II beforehand in order to allow more time for the other tasks. The cautious students who completed their reading PTs earlier spent their four weeks on three PTs. The content of all PTs was covered in three weeks and week 14 provided a chance to internalize what they had learnt. Throughout the students' implementation process instructors were expected to guide their students on PTs.

### ***Week 14***

A final contact meeting was held for the instructors in week 14 about the evaluation week. The procedures were explained and also some guiding process and product-oriented questions and feedback samples (Appendix B) were presented to the instructors. In week 14 a seminar on Learning-Oriented Assessment, lasting two

hours, was held by Professor Farhady, a lecturer at Yeditepe University and an expert on assessment, in order to raise awareness of the instructors, one week before the assessment week. The content of the seminar was predominantly theoretical but some samples of PBA and practical solutions were also presented by the guest speaker. At the same week the class hours allotted to PT presentations were highlighted in red in the original weekly programs to inform students (Appendix C) and posted on the class doors with a final warning notice including a brief explanation of week 15 (Appendix A). The attendance of the whole week would be taken in those two hours and students with PTs particularly had to attend all these class hours in order to present their tasks during the week.

#### ***Week 15 Assessment and Evaluation week***

As abovementioned for the evaluation of PTs, one week in syllabus was allotted to the evaluation and assessment of PTs as two hour in-class activity, in which students did presentations of their PTs.

In the assessment and evaluation week, the main aim of which was to give students an active role in their learning process and to foster student-student interaction as the students presented their PTs in front of the class (Appendix H). The instructors and their classmates asked some general questions about the process of preparing their PTs or some detailed questions about them. Some of the students even had the chance to perform the game or quiz show that they had prepared by implicating his or her classmates in the performance. As the presentation of PTs was not sufficient to score the final performances or products of the students, instructors spent the rest of the time to examine those in detail by using the relevant scoring rubric. The final scores were determined for each PT and were supposed to be handed in to the student affairs office until the end of week 15. Even though each instructor

has three PTs to score on average, no deadline was announced for the instructors by the administration. Therefore, the instructors were flexible to complete the scoring procedure until the official scoring deadline determined and announced as Tuesday on week 17.

### ***Week 16***

The students took their second midterm exams after completing their PT presentation process. They were supposed to study for the midterms during their spare time on week 15. They were also expected to practice and revise the exam topics through their PTs.

### **3.4.6. Assessing PTs**

Performance tasks were assessed through the scoring rubrics as abovementioned. All these rubrics were explained at contact meetings organized individually for each course and also were explained to the students before the PT practices.

Except the first PT prepared for extensive reading application, all PTs were assigned with the same deadline. An extra assessment and evaluation week was determined in the syllabus. The main aim of assessment week was to give the students a chance to present their own work by using their oral skills. By this way the students were supposed to take the responsibility of the originality of their PTs, too. When the students themselves were involved in the assessment process, it would foster the interaction. The students were supposed to present their PTs throughout the week. Two class hours for each course was allotted for the whole process. During these presentations, the students were supposed to answer some process-oriented and product-oriented questions, prepared and presented by the researcher (Appendix B) to guide the instructors. In this paper some feedback samples were also presented. As

PTs were both process and product-oriented, the presentations were not sufficient for the instructors to score the PTs. Therefore, instructors were supposed to score the product in their spare time according to the relevant scoring rubrics.

### **3.4.7. Targeted Learning Outcomes**

The essential purpose of applying PTs in this preparatory school was to encourage students to activate outside the class and reinforce what they learnt in their courses. They were expected to present an oral or written performance or product.

Through only one task the students were expected to use four language skills in an integrative way. The students had the opportunity to transform the input into an original creative output via PTs.

However, the underlying goal of PTs was to improve higher-order skills such as critical thinking or thinking strategically and to increase learner autonomy. Through a more-learner-centered assessment practice, it was also attempted to raise students' motivation to learn English. As mentioned in the "task descriptions" section above, all these tasks were presented with their related learning outcomes to the students. Some of them were role-playing, summarizing, and using specific lexical or structural items in a defined context by relating them to the real-life situations. In the assessment and evaluation week, during the presentations, students' communicative strategies, oral skills and presentation skills were also aimed to improve.

### **3.4.8. Final Evaluation Meeting with Instructors**

As described earlier, the performance-oriented application attempts were in use in the context for several semesters. From the very beginning until the end of the process instructors were aware that the first compulsory PT implementation of that year was a trial. Therefore, in accordance with the decision/accountability-oriented



program evaluation principles, the new assessment method was approached by the stakeholders as a component of the language program. At the opening meeting as well as in the training sessions of PTs, instructors were requested to examine the process so as to determine the strong and weak points of the implementation. In this respect both qualitative and quantitative data were collected from all stakeholders regarding their opinions. During the seminar week (week 17), all these aspects of the PT application were evaluated in detail. Immediately after the data analysis, the revisions based on the feedback from stakeholders were shared in a seminar held by the researcher (Appendix K). At the end of the year, the administration might reach a decision regarding PT implementation. Therefore, the scientific findings of this study would be vital to revise, reshape or totally eliminate the current implementation.

### **3.5. Data Collection**

The data for this study was collected through an open-ended questionnaire and a survey that included Likert-type as well as open-ended items. Both instruments were designed to get information from all stakeholders, namely students, instructors and administrators.

#### **3.5.1. Questionnaires with Open-ended Items**

The open-ended questionnaire included a consent form for voluntary participation, a brief explanation of the objective of the study and two open-ended questions on the positive and negative aspects of PTs (Appendix I). The open-ended questions were given to the participants before the PT implementation. These questions aimed to collect data on the expectations of all stakeholders regarding performance-based assessment before the implementation. Therefore, as preliminary data, anticipation of positive and negative aspects of PBA was required from the

participants. Three versions of the questionnaire were used: one was for the instructors. It was prepared in English due to the existence of native speakers at the preparatory school. Then it was translated into Turkish in order to gain more fruitful data when applied to the two administrators and to the students since it is their native language. Before conducting the questionnaires, expert opinion was taken for its content.

### **3.5.2. Survey with Likert-type and Open-ended Items**

Although the researcher investigated the literature and sought for the right questionnaire suitable for the objective of the study first, the questionnaires used in similar studies were predominantly designed for those contexts and lacked the program evaluation part. Therefore, a Likert-scale was developed by the researcher that was suitable for the present context with expert opinion. Three separate scales were used in the study (Appendix J). A questionnaire in English was designed for the instructors. The same scale was adapted and presented in Turkish for the administrators. The scale for students was developed in Turkish for the students. Parallelisms in terms of the number and the content of items were formed among these three surveys.

All three versions of the survey consisted of four main components; a consent form, a demographic information part, a scale including five sections (planning, application, scoring, learning outcomes, program consistency) and a part with open-ended questions. The sections of the scale were based on the sequence of evaluation 'planning, application and evaluation' as emphasized in the literature. As the PT implementation was an assessment practice, the scoring part was added and the learning outcomes section was incorporated by the researcher in order to evaluate its contributions to students. Each scale consisted of a total of 45 items. According to the

five-point Likert-scale, participants were asked to select the most suitable number that represents their opinion for each statement, namely 5 for strongly agree, 4 for agree, 3 for undecided, 2 for disagree, and 1 for strongly disagree. Three open-ended items attached to the end of the scales asked the participants their opinions and suggestion on the continuity of PT implementation (Appendix J). If they chose the first item, they thought that the PT practices should continue in the same way and they were expected to give their reasons. If they selected the second item, they thought that PTs should continue with some revisions and they were supposed to make suggestions. If they chose the third item, they believed PTs should not continue in the following years and they were expected to give their reasons.

The researcher relied on her literature readings and benefitted from the analysis of the qualitative data gathered for the present study while forming the items of the Likert-scale. While preparing her scales, she consulted experts for the content of the scale and two statisticians for the format and reliability of the items. The main goal was to collect both quantitative and qualitative data from all stakeholders through the questionnaires in order to gain a deeper insight into the focus point of the study.

After receiving approval from the experts, piloting was carried out for the instructor scale with the participation of six instructors. This was followed by an item analysis. In accordance with the item analysis, some items in Part II were eliminated to improve reliability. The parallel items in students' and administrators' questionnaires were also eliminated. According to the Alpha-Cronbach analysis, the reliability of the questionnaire designed for the instructors was 0, 81 in the pilot data and 0, 95 in the main data.

**Table 6.** Cronbach's Alpha information for instructors' scale

<b>Pilot Data</b>	
Cronbach's Alpha	,813
N of Items	45
<b>Actual Data</b>	
Cronbach's Alpha	,954
N of Items	45

Piloting for the students' questionnaire was executed with one random class of 22 students. The reliability was analyzed and it was observed that the reliability in terms of Cronbach's Alpha was 0,94 for the pilot data and it was 0,93 in the main data.

**Table 7.** Cronbach's Alpha information for students' scale

<b>Pilot Data</b>	
Cronbach's Alpha	,942
N of Items	45
<b>Actual Data</b>	
Cronbach's Alpha	,937
N of Items	45

As the number of the population and sample were equal for administrators and the number of respondents was limited to two, no piloting and also no reliability and validity check could be executed on the scale designed for them.

For the validity check, Kaiser-Mayer-Olkin (KMO) and Barlett's test (BTS) were carried out for both scales. As KMO value was high (,838) for students' scale and (,675) for instructors' scale and BTS results of both scales showed that p value was (,000) lower than ,01, the scales for students and instructors were suitable for factor analysis (Büyüköztürk, 2002). To check the construct validity of these scales, factor analysis, principle components analysis, was administered. The total variance of the five factors whose attributes were higher than 1,00 was % 53,150 for students'

scale while %60,202 for instructors' scale. In the literature, an analysis which explains %50-75 of the total variance was a valid analysis but %40 and higher can also be considered as a valid analysis (Ceyhan & Gürcan-Namlu, 2000, p. 77-93). Expert opinion was taken during this process.

Since there were statements related with student scores in the scale and instructors needed time to evaluate their students before responding to the items related with learning outcomes, the instrument was planned to be conducted in the beginning of the following semester after the participants were informed about the GPAs (General Point Average) for the fall semester. After approximately one month after the PT implementation the questionnaires were presented to all stakeholders.

### **3.6. Data Collection Procedures**

As mentioned above, two main instruments were used in this study to collect both quantitative and qualitative data; an open-ended questionnaire and a survey with Likert-type and open-ended items, from students, instructors, and administrators.

The initial open-ended questionnaire was presented to all of the instructors working at the preparatory school. It was delivered on the third week of fall semester, after the training on alternative assessment and just before the first PT training. The data collection of the open-ended questionnaire lasted for approximately two weeks. Although completing the survey normally took 15 minutes for all participants, it extended over time for the instructors so as to gather fruitful data particularly for the open-ended section. In order to ensure anonymity, the instructors were asked to hand in their responses to the secretary of the director. Therefore, the researcher had no information about the match between the questionnaires and their respondents. Meanwhile the questionnaires in Turkish were executed for the administrators and the students. Administrators completed the survey in the targeted duration.

To collect data from students, two classes were randomly selected from each language level, which made up a total of six classes. The survey, applied upon the completion of the PT implementation, was conducted in the same classes from which the qualitative data was collected. Although the research did not have an experimental design, the aim was to detect whether there was a shift in the participants' attitude toward PTs in positive or negative direction after the implementation. The data collection from students was initiated at the third week of the spring semester and lasted for one week. The students were given approximately 15 minutes to complete the survey. The participant students were reminded that it was the second phase of the same study.

### **3.7 Data Analysis**

The questions asked about the positive aspects as well as the negative aspects of the PTs. The qualitative data collected from students, instructors, and administrators separately were analyzed by means of content and thematic analysis. The researcher numbered all participants and coded their answers and then these codes were accumulated under relevant themes. The whole data was assigned into categories and the subtitles related with each other were merged under the same themes. Then the data was analyzed by an independent expert and a high rate of inter-rater reliability (80%) was attained.

The data gathered through the scales was analyzed by means of a statistical program to yield frequency, percentage results as well as mean scores.

The open-ended items in the survey were analyzed through the same content and thematic analysis method described above and a high rate of inter-rater reliability (80%) was attained.

## CHAPTER 4 FINDINGS

This study aimed to evaluate the assessment component of a language program at a preparatory school of a state university in respect of students', instructors' and administrators' opinions. In this chapter the results of data collected through open-ended questionnaires, Likert-scales and open-ended section of the Likert-scale will be presented in the order of the research questions. Firstly, the analysis of expectations of the participant stakeholders on the PT implementation prior to the actual practice will be presented. Secondly, the opinions of the participant stakeholders on the PT implementation and its phases upon the completion of the practice will be presented. Finally, the views and suggestions of the stakeholders on the continuity of the PT implementation will be exhibited.

**4.1.** What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of

- a. students?
- b. instructors?
- c. administrators?

The open-ended questionnaire was conducted to 126 students from three language levels (elementary, pre-intermediate and intermediate). Qualitative data collected through the questionnaire was analyzed through content and thematic analysis. The data were collected and analyzed in two aspects, namely positive and negative aspects of PTs. At the end of the analysis, the six main themes for all respondents were determined as: (A) Learning Aspect, (B) Language-related Aspect, (C) Affective Aspect, (D) Assessment-oriented Aspect (E) Task-specific Aspect and

(F) Application Aspect. In some cases, there were no responses for the themes. The subcategories and corresponding percentage of respondents are presented in tables.

**4.1.1.** What were the positive and negative aspects of performance assessment tasks before their implementation from the perspective of students?

According to the data, seen in Table 8, gathered from students on their positive expectations about PT implementation before the actual practice, the learning aspect theme consisted of eight subcategories. In terms of subcategory number, the density of responses accumulated under this theme at most. Students mostly perceived PT implementation positively in terms of its possible contribution to their learning. Rendering students' autonomous learning with the highest percentage (31%) displayed that students considered PT implementation as a contribution to their autonomous learning as all the tasks designed were out-of-class tasks. 24% of students, by the second highest percentage within the theme of learning, believed in the positive impact of PT implementation on their personal skills development such as social skills or time-management skills. 14% of participant students believed in the positive effect of PT implementation on their experiential learning. 13% of students predicted that PT implementation would contribute to the reinforcement of the course content since the application was content-based. 7% of students anticipated that PT implementation would embody different learning methods such as sub-conscious learning, process-oriented learning, and learning through technology. 7% of students considered PT implementation as a positive practice for the retention of their learning, which means students believed that their learning would be more permanent thanks to those tasks. 6% of students believed in the contribution of the PT implementation to their higher order skills such as critical thinking or problem-solving skills while 5% of



the participant students expected PT implementation to support their future goals in terms of their academic and business life.

There were three ideas grouped under the theme of language-related aspect. The highest percentage of 48% indicated that students believed that PT implementation would contribute to their general English proficiency. 20% of students believed in the possibility of PTs' positive impact on the improvement of some specific language skills and areas. The leading ones were mentioned as reading skills and speaking skills and lexical knowledge. 10% of students prospectively estimated that PT practices would assist their authentic language use, which means they expected the performance tasks to be authentic.

The affective aspect contained four main ideas. The highest percentage among those subcategories with the percentage of 14% indicated that students expected PT implementation to be a fun activity. 12% of students believed that PT practices would cause less stress compared to TA while 6% of students expected PT implementation to increase their self-confidence. 3% of students expected an increase in their motivation to learn English through PTs.

Assessment-oriented aspect consisted of three subcategories. 19% of students predicted that PT implementation would provide flexibility for students in various aspects such as time, correction chance and foreknown content, particularly compared to TA practices. 17% of students positively believed that PT implementation would give them an edge in terms of scoring contribution to their GPA. It displayed that students considered PT practices as an advantage for their success. 11% of students predicted that PBA as an assessment method would require less effort for students than their pen and paper exams.

Task-specific aspect contained three related ideas. 4% of students expressed their affirmation on the benefits of video-shooting. While 2% of students favoured the possibility of options in PTs, the same number of students thought that group and pair-work chances in PTs would have a positive impact.

Opinions related with application theme did not exist in students' positive expectations of PT practices.

**Table 8.** Positive Aspects of PTs as expected by students before the implementation (n=126)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>		
Rendering students' autonomous learning	39	31%
Development of students' personal skills	30	24%
Encouraging experiential learning	18	14%
Reinforcement of course content	16	13%
Embodying different learning methods	9	7%
Ensuring retention of students' learning	9	7%
Improving students' higher order skills	8	6%
Contribution to students' future goals	6	5%
<b>B. LANGUAGE-RELATED ASPECT</b>		
Improving general English proficiency	60	48%
Improving specific language skills & areas	25	20%
Encouraging more authentic language use	13	10%
<b>C. AFFECTIVE ASPECT</b>		
Enjoyable application for students	18	14%
Lower stress level for students	15	12%
Increasing students' self-confidence	8	6%
Increasing students' motivation	4	3%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>		
Providing flexibility in various aspects (Time & Correction & Content)	24	19%
Contribution to students' GPA	21	17%
Requiring less effort for students (compared to TA)	14	11%
<b>E. TASK-SPECIFIC ASPECT</b>		
Benefits of video shooting for students	5	4%
Task options & variety for students	3	2%
Individual, pair and group-work options	3	2%
<b>F. APPLICATION ASPECT</b>		
	0	0%

According to the findings of the students' negative expectations on PT implementation in Table 9, none of the students mentioned any language-related issues or learning aspect of PTs. This indicates that students all agreed that PTs would

contribute to their English language development and learning in general. Yet, they had some ideas about possible negative sides of the PT implementation.

Affective aspect of students' responses consisted of five subcategories as seen in the table. When the items were analyzed specifically, 19% of students perceived PT implementation as a stressing application. 10% of students mentioned they might have no motivation to prepare PTs if they were not interesting for them. 6% of students complained about the negative impact of PT implementation on their social activities due to its duration. 5% of students believed that they would not enjoy preparing PTs while 5% of students seemed resistant or unwilling to do PTs.

Assessment-related aspect of PTs consisted of five subcategories. The highest percentages among other sections were accumulated under the theme of assessment. 49% of students, by the highest percentage in this section, believed that as an assessment method PBA requires extra effort and difficulty for them compared to one-shot pen and paper exams. Almost half of the participant students had negative expectations about the effort they would make and its difficulty. 30% of students anticipated that PTs would take too much time. 13% of students were concerned about the scoring of their PTs in terms of subjective assessment. 8% of students considered that the main goal of PT implementation would be merely getting good scores rather than learning English. 6% of students showed a tendency in TA preference instead of PBA due to some reasons such as its duration, difficulty or their unfamiliarity.

Task-specific aspect consisted of five subcategories. Among 126 students, 17% were concerned about task content and scope. 16% of students considered video shooting practice of tasks as difficult due to technological reasons or feeling uncomfortable performing in a video. 15% of students complained about the overwhelming number of PTs, which means four separate PT practices for four

courses seemed demanding for them. 13% of students commented negatively about the facilities such as access to technology and resources and the instruments such as CDs and posters. 2% of students were concerned that they might not be creative or productive enough to prepare these tasks.

Application aspect consisted of three subcategories. 18% of students had assumptions that they would not have sufficient time to complete PTs due to short deadlines. 10% of students believed that they would need guidance during the application and feedback after the application, which displayed their expectations from instructors. 7% of students thought that PT practices were not suitable for all types of learners. They mentioned that introverts or students with weak oral skills might have difficulty in performance-based assessment.

**Table 9.** Negative Aspects of PTs as expected by students before the implementation (n = 126)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>	0	0%
<b>B. LANGUAGE-RELATED ASPECT</b>	0	0%
<b>C. AFFECTIVE ASPECT</b>		
Higher stress level for students	24	19%
Lack of motivation for students	13	10%
Negative impact on students' social life	8	6%
Not enjoyable application for students	6	5%
Unwillingness / Resistance to compulsory application	6	5%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>		
Requiring extra effort & difficulty for students	62	49%
Spending too much time to prepare PTs	38	30%
Concerns about PTs' scoring	16	13%
Considering the main goal as getting good scores	10	8%
TA Preference of Students	8	6%
<b>E. TASK-SPECIFIC ASPECT</b>		
Students' concerns about task content & scope	21	17%
Unwillingness to do video shooting	20	16%
Overwhelming number of PTs	19	15%
Students' concerns about instruments and facilities	16	13%
Requiring creativity & production for students	3	2%
<b>F. APPLICATION ASPECT</b>		
Short deadlines to complete PTs	23	18%
Need for guidance & feedback	13	10%
Not suitable application for all learners	9	7%

**4.1.2.** What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of instructors?

The instrument was also given to 60 instructors. The data were again analyzed in two phases as positive and negative aspects and in each there were the same six themes as found in students but various different subcategories. These subcategories and their corresponding percentage of respondents are presented in tables.

According to the positive assumptions of instructors on PT implementation, seen in Table 10, in terms of learning, there were seven subcategories with the highest percentages. 74% of instructors, ranked in the first place of this section, stated that PT implementation would render students' autonomous learning as the PTs were prepared out of class. 48% of instructors believed that PTs would enhance students' higher order skills. 33% of instructors expected PT practices to reinforce of the course content since those practices were content-based. 27% of instructors believed that those PT practices would render process-oriented learning, different from TA practices. 23% of instructors thought that students would acquire various ways of English learning through PTs. 22% of instructors expected PT implementation to contribute to teacher-student and student-student interaction. 10% of instructors were positive about the trigger to push students to learn more while preparing and presenting their PTs in terms of the target language.

Considering their positive perceptions on PT implementation in terms of language-related aspect, there were three subcategories. 32% of instructors thought that PT practices would increase authentic language use through authentic tasks. 15% of instructors predicted that PT practices would enhance students' communicative competence and performance. 13% of instructors had positive opinion about the

contribution of PTs on the integrated skills development of students in accordance with the nature of PBA.

Five positive ideas were accumulated under the theme of affective aspect. 35% of instructors expected PTs to increase students' motivation. 22% of instructors positively stated that PT implementation would lead to less stress for students compared to TA practices while 18% of instructors expected PTs to be enjoyable for students. 13% of instructors believed in the positive impact of those practices on students' self-confidence whereas 8% of instructors mentioned its positive effect particularly on introvert students.

Participant instructors' ideas in five subcategories were labeled as assessment-oriented aspect. 52% of instructors believed that those practices would reflect holistic picture of students' assessment for both instructors and students by measuring their actual performance. 35% of instructors anticipated that PBA practices would be more learner-centered compared to TA practices. 20% of instructors assumed that PT practices would provide a positive washback effect on students' progress and impact on teaching practices by analyzing student needs. 8% of instructors positively stated that PBA method ensured time flexibility for students compared to one-shot exams. 3% of instructors expected PT practices to be time-saving for instructors.

Task-specific and application-related opinions did not exist in instructors' comments on the positive aspect of PT practices.

**Table 10.** Positive Aspects of PTs as expected by instructors before the implementation (n = 60)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>		
Rendering students' autonomous learning	44	74%
Improving students' higher order skills	29	48%
Reinforcement of the course content	20	33%
Rendering process-oriented learning	16	27%
Encouraging different ways of learning	14	23%
Empowering interaction between T -SS & SS-SS	13	22%
Triggering students to learn more	6	10%
<b>B. LANGUAGE-RELATED ASPECT</b>		
Encouraging more authentic language use	19	32%
Improving students' communicative competence & performance	9	15%
Integrated language skills development of students	8	13%
<b>C. AFFECTIVE ASPECT</b>		
Increasing students' motivation	21	35%
Lower stress level for students	13	22%
Enjoyable application for students	11	18%
Increasing students' self-confidence	8	13%
Positive impact on introvert students	5	8%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>		
Reflecting holistic picture of student assessment	31	52%
More learner-centered assessment method	21	35%
Washback effect & impact of PBA	12	20%
Assessment method with time flexibility for students	5	8%
Time-saving assessment method for instructors	2	3%
<b>E. TASK-SPECIFIC ASPECT</b>		
	0	0%
<b>F. APPLICATION ASPECT</b>		
	0	0%

As presented in Table 11, none of the instructors expressed negative opinions on language-related and learning aspects of PT implementation.

According to the negative anticipations of instructors on PT implementation, participant instructors mentioned four subcategories related with affective aspect. 23% of instructors believed that students might not be motivated enough to prepare PTs. 10% of instructors believed in the negative effect of PTs on introvert students. 5% of instructors stated that students might feel stressed while preparing and presenting PTs compared to TA practices which they were more familiar with. 3% of instructors stated that instructors might not be motivated for the PT implementation.

The theme assessment-oriented aspect included six different ideas as subcategories. 52% of instructors, more than half of them, mentioned their concerns about assessment dimensions such as reliability, validity and practicality. 45% of instructors, almost half of the participants, had concerns about the possibility of subjectivity while scoring PTs. 23% of instructors expected PBA would require a lot of time to score for instructors while the same number of them emphasized students' time and effort to prepare PTs. 23% of instructors stated that students would cheat on PTs by stealing information from other resources. 3% of instructors believed that the time allowed for scoring PTs would not be adequate for them.

Instructor responses in terms of task-specific aspect contained five subcategories. 8% of instructors thought that preparing different PTs for four courses separately would be demanding for students plus their traditional exams. 5% of instructors predicted that they would have difficulty while grouping or pairing students for the PTs. 5% of instructors stated that limited content of PTs would have a negative effect. 3% of instructors estimated that task options would not be limited for students as it might limit their creativity while the same percentage of instructors believed that if tasks were difficult, it would affect students negatively.

Application theme consisted of five subcategories. 30% of respondents thought that PT implementation would require training and guidance for students whereas 25% of instructors assumed that instructors would need training and guidance before or during the PT practices. 10% of instructors thought PT practices might cause students to have time management problems due to short deadlines. 7% of instructors expected PTs not to be suitable for crowded classes while 3% of participants believed that students would need more feedback for correction and improvement during or after the PT implementation.



**Table 11.** Negative Aspects of PTs as expected by instructors before implementation

(n=60)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>	0	0%
<b>B. LANGUAGE-RELATED ASPECT</b>	0	0%
<b>C. AFFECTIVE ASPECT</b>		
Lack of motivation for students	14	23%
Negative impact on introvert students	6	10%
Higher stress level for students	3	5%
Lack of motivation for instructors	2	3%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>		
Concerns about assessment dimensions (e.g. reliability, validity)	31	52%
Concerns about subjectivity in assessment	27	45%
Possibility of plagiarism	14	23%
Requiring time & effort to score PTs	14	23%
Requiring too much time & effort (compared to TA)	14	23%
Limited time to assess	2	3%
<b>E. TASK-SPECIFIC ASPECT</b>		
Overwhelming number of tasks	5	8%
Difficulty of grouping & pairing students	3	5%
Limited content of PTs	3	5%
Limited task options for students	2	3%
Task difficulty for students	2	3%
<b>F. APPLICATION ASPECT</b>		
Requiring training & guidance for students	18	30%
Requiring training & guidance for instructors	15	25%
Short deadlines for students to complete PTs	6	10%
Not suitable for large-size classes	4	7%
Lack of feedback on PTs for students	2	3%

**4.1.3.** What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of administrators?

Finally, the instrument was given to two administrators. The data were collected and analyzed in two main aspects as positive and negative. The same six themes as found in students' and instructors' data and their distinct subcategories with corresponding percentages of respondents are presented in tables.

As seen in Table 12, according to the positive expectations of administrators on PT implementation in terms of learning aspect, there were four subcategories. Both

administrators anticipated that PT practices would render autonomous learning of students through those out-of-class tasks. One of the administrators expected PTs to encourage experiential learning of students. 50% of administrators believed that PT practices would improve students' higher order skills, particularly critical thinking skill. 50% of administrators believed that PTs would affect student-student and teacher-student interaction positively. The expectations of administrators more densely accumulated under the theme of learning.

There were only two subcategories under the theme of language-related aspect. One of the administrators believed in the positive effect of PT implementation on students' integrated language skills development while one of the administrators stated that PTs would have a positive impact on authentic language use.

Considering the affective aspect of PTs, administrators highlighted two expectations. Both of the administrators agreed on the contribution of PT implementation to student motivation. However, only one of them believed that PTs would likely affect students' self-confidence positively.

There was only one positive opinion related with the assessment aspect of PTs. One of the administrators anticipated that PTs would contribute to learning as it was both process and product-oriented assessment method compared to TA practices.

Neither of the administrators commented positively on PT implementation in terms of task-specific issues and issues related with application.

**Table 12.** Positive Aspects of PTs as expected by administrators before the implementation (n=2)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>		
Rendering students' autonomous learning	2	100%
Encouraging experiential learning	1	50%
Improving students' higher order skills	1	50%
Empowering SS-SS & T-SS interaction	1	50%
<b>B. LANGUAGE-RELATED ASPECT</b>		
Integrated language skills development of students	1	50%
Encouraging more authentic language use	1	50%
<b>C. AFFECTIVE ASPECT</b>		
Increasing students' motivation	2	100%
Increasing students' self-confidence	1	50%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>		
Process & Product-oriented assessment method	1	50%
<b>E. TASK-SPECIFIC ASPECT</b>		
	0	0%
<b>F. APPLICATION ASPECT</b>		
	0	0%

As presented in Table 13, neither of the administrators mentioned the learning, language-related, affective and assessment aspects of the PT implementation.

Considering administrators' negative expectations in terms of task-specific aspect before the PT practice, this theme consisted of two sub-categories. One of the administrators believed that there was a prospect for instructors to have technical deficits during or after the PT practices while one of the administrators stated that the materials used to prepare PTs would lead to waste of paper.

As a negative anticipation of administrators related with the application aspect of PTs, there was only one common idea. Both administrators expected these practices to be an extra workload for the staff.

**Table 13.** Negative Aspects of PTs as expected by administrators before the implementation (n=2)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>	0	0%
<b>B. LANGUAGE-RELATED ASPECT</b>	0	0%
<b>C. AFFECTIVE ASPECT</b>	0	0%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>	0	0%
<b>E. TASK-SPECIFIC ASPECT</b>		
Possibility of technical deficits of instructors	1	50%
Causing paper waste	1	50%
<b>F. APPLICATION ASPECT</b>		
Bringing extra workload for instructors	2	100%

- 4.2.** How were the planning, application, scoring, learning outcomes and program consistency phases of performance-based assessment tasks evaluated after their implementation by
- students?
  - instructors?
  - administrators?

The quantitative data collected from the stakeholders via the Likert-scale consisting of 45 statements were analyzed by means of a statistical program. The average of the scale as well as the averages of the subscales was analyzed through descriptive and inferential statistics. All the mean values of the reverse items were labeled with an asterisk (\*) and their original mean scores were presented in the tables below.

- 4.2.1.** How were the planning, application, scoring, learning outcomes and program consistency phases of PTs evaluated after their implementation by students?

In order to answer the second research question, opinions of 129 students on the five main phases of Performance-based assessment were collected through a scale

with 5 point Likert type items. These phases consisted of planning, application, scoring, learning outcomes and program consistency.

As seen in Table 14, whole scale results show that students generally evaluated the current PT implementation positively with an overall mean 3,68. The highest mean score ( $\bar{X}$ = 4.12) was observed for the planning phase of the PT implementation. It was followed by the application phase ( $\bar{X}$ = 3.79), the scoring phase ( $\bar{X}$ = 3.76) and the program consistency ( $\bar{X}$ = 3.40) of the PT practices. The lowest mean score belonged to learning outcomes ( $\bar{X}$ = 3.35). Even though the learning outcomes phase was the least favoured, students were between undecided and positive.

**Table 14.** Total mean values of students' opinions on the phases of PT implementation

<b>MEAN VALUES PER SUBSCALE</b>	<b>N</b>	<b><math>\bar{X}</math></b>	<b>SD</b>
Planning	129	4,12	,80
Application	129	3,79	,81
Scoring	129	3,76	1,09
Program Consistency	129	3,40	,75
Learning Outcomes	129	3,35	1,26
<b>TOTAL</b>	<b>129</b>	<b>3,68</b>	<b>,94</b>

Items one through eight were asked to obtain participants' opinions on the planning phase of the PT implementation. As seen in Table 15, in total, the participants reported to be content with the planning phase of the PT practices ( $\bar{X}$  = 4.12). Planning items received the highest total mean score of the overall questionnaire. The mean value of the planning section indicated that students were mostly satisfied with the effectiveness of the planning phase of the PT implementation. It revealed that most of the participants agreed that their instructors had sufficient knowledge on PTs and provided them with the necessary and clear information such as instructions, explanations on how to apply PTs and the scoring

rubrics before the actual practice. Students expressed strong agreement with the 7<sup>th</sup> item ( $\bar{X} = 4.64$ ) and 6<sup>th</sup> item ( $\bar{X} = 4.50$ ), which got the highest mean values, revealing the fact that students were announced the deadlines and the weekly schedules of the PTs. In addition students stated that they were satisfied with the training provided before the practices in the 2nd item with a high mean score ( $\bar{X} = 4.44$ ). However the lowest mean score in item 8 ( $\bar{X} = 2.83$ ) indicated that students were not sure whether they were ready or not for the PTs before the implementation. In accordance with the students' general positive attitude on planning stage, their disagreement with the statement in item 4\* ( $\bar{X} = 2.30$ ) displayed that most of them could understand how to prepare their PTs properly.

**Table 15.** Opinions of students on the “planning” phase after the PT implementation

Items	N	$\bar{X}$	SD
1. Instructors had sufficient information about PT implementation.	129	4,34	,61
2. Sufficient training was provided to us before the PT implementation.	129	4,44	,64
3. Instructors had sufficient information about scoring rubrics.	129	4,31	,75
4. I did not understand how to prepare PTs.*	129	2,30*	1,06
5. Clear instructions about each PT were provided to us by instructors.	129	4,17	,82
6. A weekly schedule was announced for each PT.	129	4,50	,81
7. The deadlines to hand in PTs were announced to us.	129	4,64	,54
8. As a student, I felt ready for PTs before the PT implementation.	129	2,83	1,16
<b>TOTAL</b>	<b>129</b>	<b>4,12</b>	<b>,80</b>

\*Reverse item

Items nine through eighteen were asked to obtain participants' opinions on the application phase of PTs. As seen in Table 16, in total, the participants reported to be mostly satisfied with the application phase of the PT practices ( $\bar{X} = 3.79$ ). Therefore, they generally seemed to have positive opinions about the overall application phase of the PTs. Students mostly agreed with item 17 ( $\bar{X} = 4.47$ ), item 18 ( $\bar{X} = 4.43$ ), item 9 ( $\bar{X} = 3.81$ ) and item 10 ( $\bar{X} = 3.81$ ), which got the highest mean values. When these values were interpreted respectively, the students claimed that they managed to hand in their PTs on time. They indicated that they prepared their PTs themselves without getting help from others. They also agreed that PTs suited their cognitive and

language levels. The mean value in item 13\* ( $\bar{X} = 3.26$ ) displayed that students had some doubts whether to consider PTs as a time-consuming practice or not. Although they showed indecisiveness about the joy they got while doing PTs in item 11, the mean score of 15<sup>th</sup> item\* ( $\bar{X} = 3.67$ ) indicated that they slightly agreed that they got bored while preparing them. The lowest means of this section in item 16 ( $\bar{X} = 2.77$ ) and item 11 ( $\bar{X} = 2.60$ ) revealed the fact that they were not sure about the effectiveness of presenting PTs in the class and about the joy they got from PTs.

**Table 16.** Opinions of students on the “application” phase after the PT implementation

Items	N	$\bar{X}$	SD
9. PTs were suitable for my language level.	129	3,81	,61
10. PTs were suitable for my cognitive level.	129	3,81	,64
11. PTs were enjoyable.	129	2,60	,75
12. Sufficient time was given to us for the completion of PTs.	129	3,55	1,06
13. PT implementation was time-consuming.*	129	3,26*	,82
14. Sufficient guidance was provided to us during PT process.	129	3,56	,81
15. I got bored while doing PTs.*	129	3,67*	,54
16. Presenting PTs in class was effective for me.	129	2,77	1,16
17. I could hand in my PTs on time.	129	4,47	,91
18. I prepared my PTs on my own.	129	4,43	,87
<b>TOTAL</b>	<b>129</b>	<b>3,79</b>	<b>,81</b>

\*Reverse item

Items nineteen through twenty-six were asked to obtain participants’ opinions on the scoring phase of PTs. When the specific items were analysed as presented in Table 17, the total mean value ( $\bar{X} = 3.76$ ) indicated that the participant students were mostly positive about the scoring phase of these practices. Most of the students agreed with item 19 ( $\bar{X} = 4.00$ ), item 23 ( $\bar{X} = 3.94$ ), and item 26 ( $\bar{X} = 3.91$ ). These findings revealed that most of the students agreed with scoring criteria usage. They believed that the duration allotted to scoring was sufficient. They also agreed that not only the process but also the product was assessed as planned. Considering item 25, which got the second lowest mean value ( $\bar{X} = 3.37$ ), it was found out that respondents were not sure about the feedback given for the PTs. The students mostly disagreed with item

24\* ( $\bar{X}$  = 2.19) related with the unfair scoring in consistency with a great number of students' agreement on objective scoring of PTs in item 21 ( $\bar{X}$  = 3.89).

**Table 17.** Opinions of students on the “scoring” phase after the PT implementation

Items	N	$\bar{X}$	SD
19. Each PT was evaluated through a rubric.	129	4,00	,95
20. Scoring rubrics were easy to understand.	129	3,67	1,10
21. PTs were scored objectively.	129	3,89	1,13
22. PT scores were satisfying.	129	3,47	1,21
23. Sufficient time (1 week) was allotted for assessment.	129	3,94	1,01
24. PTs were scored unfairly.*	129	2,19*	1,11
25. Feedback about our PTs was provided to us.	129	3,37	1,30
26. Both process and product of PTs were scored as planned.	129	3,91	,91
<b>TOTAL</b>	<b>129</b>	<b>3,76</b>	<b>1,09</b>

\*Reverse item

In order to obtain participants' opinions on the learning outcomes phase of PTs, items twenty-seven through thirty-eight were asked to the participants. Findings presented in Table 18 showed that in total the participants reported to be between undecided and agree with the learning outcomes of the PT practices ( $\bar{X}$  = 3.35). This subscale had the lowest average of the overall questionnaire. Within this subscale, students mostly agreed with the item 32 ( $\bar{X}$  = 4.05), item 28 ( $\bar{X}$  = 3.67), item 33 ( $\bar{X}$  = 3.66), which got the highest mean values. According to the findings deduced from the items that were agreed by students, it was indicated that they believed in the contributions of PTs to their lexical knowledge, reading skills, general proficiency in English respectively. The items with low mean scores were item 30 ( $\bar{X}$  = 2.93), item 36 ( $\bar{X}$  = 2.81) and item 35 ( $\bar{X}$  = 2.53), revealing that they were quite not sure about PTs' contribution to their listening skills, their confidence levels and their motivation. Students disagreed with item 37\* ( $\bar{X}$  = 2.48) and highlighted that they believed in the contribution of PTs to their general English ability.



**Table 18.** Opinions of students on the “learning outcomes” phase after PT implementation

Items	N	$\bar{X}$	SD
27. PTs contributed to the development of my writing skills.	129	3,39	1,13
28. PTs contributed to the development of my reading skills.	129	3,67	1,08
29. PTs contributed to the development of my speaking skills.	129	3,56	2,04
30. PTs contributed to the development of my listening skills.	129	2,93	1,17
31. PTs contributed to the development of my grammatical knowledge.	129	3,37	1,15
32. PTs contributed to the development of my vocabulary knowledge.	129	4,05	1,87
33. PTs contributed to the development of my general language proficiency.	129	3,66	,99
34. PTs increased my autonomy.	129	3,33	1,08
35. PTs increased my motivation to learn English.	129	2,53	1,15
36. PTs increased my self-confidence.	129	2,81	1,20
37. I do not think that PTs contributed to my general language proficiency.*	129	2,48*	1,14
38. I reinforced what I have learnt in courses through PTs.	129	3,39	1,13
<b>TOTAL</b>	<b>129</b>	<b>3,35</b>	<b>1,26</b>

\*Reverse item

In order to obtain participants’ opinions on the program consistency phase of PTs, items thirty-nine through forty-five were asked to the participants. As seen in Table 19, in total, the participants reported to be undecided about the program consistency of the PT practices ( $\bar{X} = 3.40$ ) with a mean value close to agree. The 41<sup>st</sup> item ( $\bar{X} = 3.86$ ) received the highest mean rating followed by the 43<sup>rd</sup> item ( $\bar{X} = 3.67$ ). The students mostly believed that PTs were consistent with the content of courses and with the principles of PBA. Low mean scores were seen in item 40 ( $\bar{X} = 2.91$ ) and item 39 ( $\bar{X} = 2.86$ ). This indicated that participants were unsure about the effectiveness of PBA individually or its superiority to TA. They seemed to have no precise opinions about this kind of comparison or any dominant tendency. However much higher mean score in item 45 ( $\bar{X} = 3.34$ ), very close to agreement yet still doubtfulness showed that students seemed to consider PBA as a supplementary tool. In parallelism with the mean value of item 41, the respondents’ disagreement with the 44<sup>th</sup> item\* ( $\bar{X} = 2.22$ ) revealed the fact that PTs had a content-validity.

**Table 19.** Opinions of students on the “program consistency” after the PT implementation

Items	N	$\bar{X}$	SD
39. Performance-based assessment is an effective assessment method for ELT.	129	2,86	,61
40. Performance-based assessment methods are more useful than traditional exams for language learning and teaching.	129	2,91	,64
41. PT implementation applied in this institution was consistent with the content of the courses.	129	3,86	,75
42. Performance-based assessment elements were in accordance with the objectives of the language program of this prep-school.	129	3,35	,82
43. PT implementation in this institution reflected the principles of performance-based assessment.	129	3,67	1,06
44. PT implementation conducted in this institution did not reflect what we have learnt. *	129	2,22*	,81
45. Performance-based assessment is useful only as a supplementary tool to traditional assessment.	129	3,34	,54
<b>TOTAL</b>	<b>129</b>	<b>3,40</b>	<b>,75</b>

\*Reverse item

#### 4.2.2. How were the planning, application, scoring, learning outcomes and program consistency phases of PTs evaluated after their implementation by instructors?

In order to answer the second research question, opinions of 60 instructors on the five main phases of PT implementation were collected as well. The mean scores for each of these parts of the scale as well as the subscales are presented in the tables.

According to these findings in Table 20, the total mean value of the questionnaire for instructors which was  $\bar{X}=3.89$  indicated that instructors slightly more than moderately agreed on all five sections of the Likert-scale, with the highest score in planning ( $\bar{X} = 4.18$ ). Instructors mostly had highly positive opinions on the current PT implementation and more positive compared to the students. The second highest mean score ( $\bar{X} = 4.03$ ) was the scoring phase followed by the program consistency phase ( $\bar{X} = 3.84$ ). While the application phase got a mean value of 3.81, the lowest mean value ( $\bar{X} = 3.61$ ) among these phases belonged to the learning outcomes of the PT practices in terms of instructor opinions.

**Table 20.** Total mean values of instructors' opinions on the phases of PT implementation

<b>MEAN VALUES PER SUBSCALE</b>	<b>N</b>	<b><math>\bar{X}</math></b>	<b>SD</b>
Planning	60	4,18	,78
Scoring	60	4,03	,83
Program Consistency	60	3,84	,75
Application	60	3,81	,76
Learning Outcomes	60	3,61	,83
<b>TOTAL</b>	<b>60</b>	<b>3,89</b>	<b>,79</b>

Items one through eight were asked to obtain instructors' opinions on the planning phase of the PT implementation. As presented in Table 21, in total, the participants mostly agreed that the planning phase of the PT practices was carried out properly ( $\bar{X} = 4.18$ ) with the highest mean value of the whole questionnaire. This phase was more appreciated than others. Instructors mostly agreed with the 7<sup>th</sup> item ( $\bar{X} = 4.65$ ) and 6<sup>th</sup> item ( $\bar{X} = 4.55$ ), which got the highest values, revealing that they adopted a strong claim that students handed in their PTs on time and a weekly schedule was announced. The other item with a high mean score was the 3<sup>rd</sup> item ( $\bar{X} = 4.37$ ). This agreement in item 3 indicated that majority of instructors believed that they were provided with explanations of scoring rubrics before the practices. The second lowest mean value in item 1 ( $\bar{X} = 3.78$ ) displayed that respondents were not sure about the adequacy of the training they got before the PT practices. The lowest mean value ( $\bar{X} = 2.17$ ) of item 4\* revealed that participants mostly believed in students' comprehension of PT requirements in consistency with their agreement on clear PT instructions and explanations.

**Table 21.** Opinions of instructors on the “planning” phase after the PT implementation

Items	N	$\bar{X}$	SD
1. Sufficient training was provided to me before the PT implementation.	60	3,78	1,01
2. Sufficient training was provided to students before the PT implementation.	60	3,85	1,04
3. Rubrics were explained to me before the PT implementation.	60	4,37	,61
4. Students did not understand the requirements of the PTs.*	60	2,17*	,81
5. Students were given clear instructions about each PT by instructors.	60	4,22	,74
6. A weekly schedule was announced for each PT.	60	4,55	,65
7. The deadlines to hand in PTs were announced to the students.	60	4,65	,55
8. Students were well-informed about the whole process before PT implementation started.	60	4,18	,79
<b>TOTAL</b>	<b>60</b>	<b>4,18</b>	<b>,78</b>

\*Reverse item

Instructors responded to the items nine through eighteen to give their opinions on the application phase of the PTs. As seen in Table 22, in total most of the participants agreed that the application phase of the PT practices were held effectively ( $\bar{X} = 3.81$ ), which means generally they were satisfied with the way of putting PTs into practice. The item with the highest mean score ( $\bar{X} = 4.38$ ) was item 12, revealing that the time allotted to PTs was enough for the students to complete them on time. That highest score was followed by item 17 ( $\bar{X} = 4.27$ ), revealing that most of the respondents believed that students could deliver their PTs before the deadlines. Item 9 and item 10 which got the same mean value ( $\bar{X} = 4.10$ ) indicated that instructors mostly agreed on PTs’ suitability for both students’ cognitive and language level. In item 15\* ( $\bar{X} = 2.88$ ) and item 13\* ( $\bar{X} = 2.90$ ), which got the lowest values, it was found out that instructors mostly did not have strong ideas about whether PT application was an enjoyable activity for the students or not. Majority of them were also not sure if it was a time-consuming activity for the instructors themselves or not.

**Table 22.** Opinions of instructors on the” application” phase after the PT implementation

Items	N	$\bar{X}$	SD
9. PTs were suitable for students’ language level.	60	4,10	,75
10. PTs were suitable for students’ cognitive level.	60	4,10	,82
11. PTs were enjoyable for students.	60	3,50	,87
12. Sufficient time was given to students for the completion of PTs.	60	4,38	,80
13. PT implementation was time-consuming for me as an instructor.*	60	2,90*	1,20
14. Sufficient guidance was provided to me during the PT process.	60	4,08	,83
15. Students got bored while preparing PTs.*	60	2,88*	,85
16. Presenting PTs in class was an effective part of the implementation for students.	60	3,65	,82
17. Students were able to hand in PTs on time.	60	4,27	,71
18. Most students prepared their PTs on their own.	60	3,70	,81
<b>TOTAL</b>	<b>60</b>	<b>3,81</b>	<b>,76</b>

\*Reverse item

The items from nineteen to twenty-six exhibited the instructors’ opinions on the scoring phase of the PT practices. As seen in Table 23, in total, the participants mostly appreciated the scoring phase of these practices ( $\bar{X} = 4.03$ ). Despite the confusion on scoring objectively or not, they seemed to have positive opinions on the assessment phase of the PTs. The highest score was seen in the 19<sup>th</sup> item ( $\bar{X} = 4.40$ ) which revealed that instructors used scoring rubrics to assess each PT. The high mean score in item 23 ( $\bar{X} = 4.30$ ) indicated that most of the participants agreed that they had adequate time to assess PTs. According to the mean score of the 26<sup>th</sup> item ( $\bar{X} = 4.10$ ), instructors mostly believed that they assessed both the process and product of PTs as planned before the actual practices. The second lowest mean value of this section in the 20<sup>th</sup> item ( $\bar{X} = 3.80$ ) revealed the fact that most of the respondents could easily follow the relevant rubrics while scoring PTs. As most participants agreed on their objective scoring in item 21 ( $\bar{X} = 3.88$ ), the disagreement with item 24\* ( $\bar{X} = 2.17$ ) stressed that instructors mostly thought that they were able to assess PTs fairly.

**Table 23.** Opinions of instructors on the “scoring” phase after the PT implementation

Items	N	$\bar{X}$	SD
19. Each PT was evaluated through a rubric.	60	4,40	,72
20. Scoring rubrics were easy to follow for instructors.	60	3,80	,95
21. I could score PTs objectively.	60	3,88	,83
22. PT scores were satisfying for most of the students.	60	3,92	,77
23. Sufficient time was allotted to me for scoring.	60	4,30	,81
24. I could not score PTs fairly.*	60	2,17*	,81
25. I gave feedback to most of the students about their PTs.	60	4,00	,88
26. I scored both process and product as planned.	60	4,10	,90
<b>TOTAL</b>	<b>60</b>	<b>4,03</b>	<b>,83</b>

\*Reverse item

The items from twenty-seven to thirty-eight were asked to respondents to show their level of agreement on the learning outcomes of the PT practices. As seen in Table 24, the total mean value ( $\bar{X} = 3.61$ ) indicated that the participants agreed that students’ learning was improved in some aspects through the PT practices. However, this mean value was the lowest score among other sections of the scale. They stated that a number of learning outcomes could emerge as a result of PT practices. The highest mean score ( $\bar{X} = 3.82$ ) in item 38 indicated that instructors generally agreed that students reinforced the course content through PT practices. The mean value of the 32<sup>nd</sup> item ( $\bar{X} = 3.80$ ) showed that participants mostly believed in the contribution of PTs on students’ lexical knowledge. Instructors also mostly agreed that PTs had a positive impact on the students’ general English proficiency as seen in item 33 ( $\bar{X} = 3.77$ ) in consistency with their disagreement shown in item 37\* ( $\bar{X} = 2.25$ ). According to the second and third lowest mean values of this section observed in item 30 ( $\bar{X} = 3.32$ ) and item 35 ( $\bar{X} = 3.30$ ), respondents mostly had some doubts about the positive effect of PTs on students’ listening skill development and their motivation to learn English.

**Table 24.** Opinions of instructors on “learning outcomes” phase after the PT implementation

Items	N	$\bar{X}$	SD
27. PTs contributed to the development of students’ writing skills.	60	3,68	,77
28. PTs contributed to the development of students’ reading skills.	60	3,58	,81
29. PTs contributed to the development of students’ speaking skills.	60	3,72	,96
30. PTs contributed to the development of students’ listening skills.	60	3,32	,87
31. PTs contributed to the development of students’ grammatical knowledge.	60	3,47	,79
32. PTs contributed to the development of students’ vocabulary knowledge.	60	3,80	,75
33. PTs contributed to the development of students’ general language proficiency.	60	3,77	,77
34. Students’ autonomy seemed to increase through PTs.	60	3,52	,89
35. Students’ motivation to learn English was affected positively through PTs.	60	3,30	,87
36. PTs had a positive impact on students’ self-confidence.	60	3,58	,79
37. PTs did not contribute to improve students’ general language proficiency.*	60	2,25*	,95
38. PTs helped students reinforce what they have learnt in the courses.	60	3,82	,75
<b>TOTAL</b>	<b>60</b>	<b>3,61</b>	<b>,83</b>

\*Reverse item

Participant instructors answered the items from thirty-nine to forty-five in order to evaluate PTs’ consistency with the current language program carried out in that institution. According to the findings presented in Table 25, the total mean score ( $\bar{X} = 3.84$ ) indicated that the participants mostly believed in the program’s harmony with the PT practices. When specific items were analysed, it was found out that they generally presented positive opinions on the consistency of the current PT implementation with the language program. The highest mean score ( $\bar{X} = 4.10$ ) in item 41 revealed that they agreed on the content-validity of PT implementation as also justified with the lowest mean score ( $\bar{X} = 2.92$ ) in item 44\*. It still seemed to be between agreeing and being indecisive in this issue. Item 41 was followed by item 43 ( $\bar{X} = 4.03$ ) indicating that most of the respondents believed that PTs showed an agreement with the objectives of the current language program. According to the second and third lowest mean values in item 45 ( $\bar{X} = 3.57$ ) and item 40 ( $\bar{X} = 3.47$ ), while instructors mostly agreed with the effectiveness of PBA as a supplementary tool

to TA, they slightly agreed with its effectiveness compared to TA in terms of language learning and teaching. They seemed to favour the combination of both methods.

**Table 25.** Opinions of instructors on “program consistency” phase after the PT implementation

<b>Items</b>	<b>N</b>	<b><math>\bar{X}</math></b>	<b>SD</b>
39. Performance-based assessment is an effective assessment method for ELT.	60	3,92	,61
40. Performance-based assessment methods are more useful than traditional exams for language learning and teaching.	60	3,47	,64
41. PT implementation applied in this institution was consistent with the content of the courses.	60	4,10	,75
42. PT implementation in this institution reflected the principles of performance-based assessment.	60	3,75	1,10
43. Performance-based assessment elements were in accordance with the objectives of the language program of this prep-school.	60	4,03	,80
44. PT implementation conducted in this institution did not reflect what students have learnt. *	60	2,92*	,81
45. Performance-based assessment is useful only as a supplementary tool to traditional assessment.	60	3,57	,54
<b>TOTAL</b>	<b>60</b>	<b>3,84</b>	<b>,75</b>

\*Reverse item

#### 4.2.3 How were the planning, application, scoring, learning outcomes and program consistency phases of PTs evaluated after their implementation by administrators?

Opinions of two administrators on the five main phases of PT implementation were also gathered. The mean scores for the scale as well as the subscale are presented in the tables below.

According to the mean values per subscale in the questionnaire executed to administrators, the total mean value ( $\bar{X} = 4.69$ ) displayed that administrators had very positive opinions on the PT implementation in its all aspects and phases. As presented in Table 26, the findings indicated that both administrators who participated in the survey strongly agreed with all sections with the highest score being in scoring



section ( $\bar{X} = 4.87$ ) a different result compared to the students' and instructors' highest mean scores. Moreover, planning and learning outcomes phases got the same mean value ( $\bar{X} = 4.75$ ) whereas they were followed by the mean value ( $\bar{X} = 4.60$ ) of the application phase. The lowest score of the scale was seen in the program consistency phase ( $\bar{X} = 4.50$ ). This mean score revealed that both participants had a strong agreement even with the section having the lowest mean value.

**Table 26.** Total mean values of instructors' opinions on the phases of PT implementation

<b>MEAN VALUES PER SUBSCALE</b>	<b>N</b>	<b><math>\bar{X}</math></b>	<b>SD</b>
Scoring	2	4.87	,18
Planning	2	4,75	,18
Learning Outcomes	2	4,75	,36
Application	2	4.60	,28
Program Consistency	2	4.50	,51
<b>TOTAL</b>	<b>2</b>	<b>4,69</b>	<b>,30</b>

The administrators responded to the items one through eight in order to evaluate the planning phase of the PT implementation. As presented in Table 27, the total mean value ( $\bar{X} = 4.75$ ) indicated that both participants strongly agreed that the planning phase of the PT practices was carried out properly. The highest scores ( $\bar{X} = 5.00$ ) in items 1, 2, 3, 6, and 7 indicated respectively that both respondents totally agreed that sufficient training was provided to both instructors and students, rubrics were explained to those and the announcement of PT weekly schedules and deadlines were made. The lowest mean score seen in item 4\* ( $\bar{X} = 1.50$ ) revealed that both participants also believed that how to prepare PTs was clearly understood by the students. The mean score in item 5 ( $\bar{X} = 4.00$ ) indicated that both respondents agreed that explicit instructions were given to students before preparing the PTs.

**Table 27.** Opinions of administrators on “planning” phase after the PT implementation

Items	N	$\bar{X}$	SD
1. Sufficient training was provided to instructors before PT implementation.	2	5,00	,00
2. Sufficient training was provided to students before PT implementation.	2	5,00	,00
3. Scoring rubrics were explained to the stakeholders before PT implementation.	2	5,00	,00
4. Students did not understand how to do PTs.*	2	1,50*	,71
5. PT practice instructions were clearly explained to students.	2	4,00	,00
6. PT weekly schedule was announced.	2	5,00	,00
7. PT deadlines were announced to students.	2	5,00	,00
8. Students were rendered to be ready for PTs.	2	4,50	,71
<b>TOTAL</b>	<b>2</b>	<b>4,75</b>	<b>,18</b>

\*Reverse item

The administrators responded the items nine through eighteen in order to give their opinions on the application phase of the PT practices. As seen in Table 28, in total ( $\bar{X} = 4.60$ ), both of the participants strongly agreed that PT practices were put into practice properly. Both administrators strongly agreed with the 9<sup>th</sup>, 10<sup>th</sup>, 12<sup>th</sup>, and 16<sup>th</sup> items with the highest scores ( $\bar{X} = 5.00$ ), indicating respectively that PTs were in accordance with students’ cognitive and language levels, students had enough time to complete PTs and it was useful for students to present PTs in class. The moderately low score ( $\bar{X} = 3.50$ ) in item 14 displayed that both participants agreed that students were provided with adequate training while they were preparing their PTs. This score also showed that participant administrators were slightly undecided about the adequacy of the training provided for the students. The second lowest mean score in item 15\* ( $\bar{X} = 1.50$ ) revealed the fact that both respondents had a claim that students did not get bored while preparing PTs. The lowest mean score belonged to item 13\* ( $\bar{X} = 1.00$ ), which means both administrators strongly believed that PT application process was not time-consuming for the participants.

**Table 28.** Opinions of administrators on “application” phase after the PT implementation

Items	N	$\bar{X}$	SD
9. PTs were suitable for students’ language level.	2	5,00	,00
10. PTs were suitable for students’ cognitive level.	2	5,00	,00
11. PTs were enjoyable for students.	2	4,50	,71
12. The time allotted to complete PTs was sufficient.	2	5,00	,00
13. PTs were time-consuming for stakeholders.*	2	1,00*	,00
14. Sufficient guidance was provided to students during PT process.	2	3,50	,71
15. Students felt bored while doing PTs*	2	1,50*	,71
16. Presenting PTs in class was an effective part of the implementation for students.	2	5,00	,00
17. Students were able to hand in PTs on time.	2	4,00	,00
18. Most of the students prepared PTs on their own.	2	4,50	,71
<b>TOTAL</b>	<b>2</b>	<b>4,60</b>	<b>,28</b>

\*Reverse item

Item nineteen through twenty-six were asked to respondents in order to obtain their opinions on the scoring phase of the PT implementation. According to the findings presented in Table 29, both participants strongly agreed with the scoring phase of the PT practices ( $\bar{X} = 4.78$ ). Both of them strongly agreed with items 19, 21, 23, 25, and 26 with the highest score 5.00 of the Likert-scale except two items. Analysing items respectively, this strong agreement revealed the fact that participants strongly claimed that instructors assessed PTs by the means of scoring rubrics, this scoring was objective and allowing one week for the assessment was enough. They also strongly believed that PT feedback was provided by the instructors and instructors assessed both the process and the product of PTs as planned before the actual practice. Their strong agreement with item 20 ( $\bar{X} = 4.50$ ) indicated that administrators did not think scoring rubrics were easy to understand for stakeholders. According to the other lowest mean score ( $\bar{X} = 4.50$ ) in item 22, it was found out that both respondents strongly believed that most students were satisfied with their PT scores. The lowest mean score in item 24\* ( $\bar{X} = 1.00$ ) indicated that both

administrators thought that PTs were assessed by the instructors in a fair way in accordance with their strong claim in item 21.

**Table 29.** Opinions of administrators on “scoring” phase after the PT implementation

Items	N	$\bar{X}$	SD
19. Each PT was evaluated through a rubric.	2	5,00	,00
20. Scoring rubrics were easy to follow for stakeholders.	2	4,50	,71
21. PTs were scored objectively.	2	5,00	,00
22. PT scores were satisfying for most of the students.	2	4,50	,71
23. Sufficient time (1 week) was allotted for scoring.	2	5,00	,00
24. PTs were scored unfairly. *	2	1,00*	,00
25. Students were given feedback about their PTs.	2	5,00	,00
26. Both the process and product of PTs were scored as planned.	2	5,00	,00
<b>TOTAL</b>	<b>2</b>	<b>4.87</b>	<b>,18</b>

\*Reverse item

Item twenty-seven through thirty-eight were responded by both administrators to obtain their opinions on the learning outcomes phase of the PT implementation. As seen in Table 30, in total the participants strongly agreed with the learning outcomes of the PT practices ( $\bar{X} = 4.75$ ). Administrators strongly agreed with items 33, 34, 35, 36, and 38 by the highest mean score 5.00. Interpreting respectively, both administrators strongly claimed that PTs had a positive impact on students’ general language proficiency, learner-autonomy, their motivation, self-confidence and the reinforcement of the course content. The mean score ( $\bar{X} = 4.50$ ) seen in items 27, 28, 29, 30, 31, and 32 indicated that administrators strongly agreed with these items. Interpreting respectively, they strongly believed in the contribution of PTs on students’ writing, reading and speaking skills. They also strongly agreed with PTs’ positive impact on students’ grammatical and lexical knowledge. The lowest mean value seen in the 37<sup>th</sup> item\* ( $\bar{X} = 1.00$ ) highlighted that they strongly agreed that students’ general English proficiency was affected quite positively through the current PT implementation.

**Table 30.** Opinions of administrators on “learning outcomes” phase after the PT implementation

Items	N	$\bar{X}$	SD
27. PTs contributed to the development of students’ writing skills.	2	4,50	,71
28. PTs contributed to the development of students’ reading skills.	2	4,50	,71
29. PTs contributed to the development of students’ speaking skills.	2	4,50	,71
30. PTs contributed to the development of students’ listening skills.	2	4,50	,71
31. PTs contributed to the development of students’ grammatical knowledge.	2	4,50	,71
32. PTs contributed to the development of students’ vocabulary knowledge.	2	4,50	,71
33. PTs contributed to the development of students’ general language proficiency.	2	5,00	,00
34. Students’ autonomy seemed to increase through PTs.	2	5,00	,00
35. Students’ motivation to learn English was affected positively through PTs.	2	5,00	,00
36. PTs had a positive impact on students’ self-confidence.	2	5,00	,00
37. PTs did not contribute to improve students’ general language proficiency.*	2	1,00*	,00
38. PTs helped students reinforce what they have learnt in the courses.	2	5,00	,00
<b>TOTAL</b>	<b>2</b>	<b>4,75</b>	<b>,36</b>

\*Reverse item

As presented in Table 31, item thirty-nine through forty-five were asked to administrators to gather their opinions on the program consistency phase of the PT implementation. In total, the participants strongly agreed that PTs and the current language program were in consistency ( $\bar{X} = 4.50$ ). The highest mean value ( $\bar{X} = 5.00$ ) in the 42<sup>nd</sup> and the 43<sup>rd</sup> items revealed that PT implementation carried out in that preparatory school was in accordance with its program objectives and with the main principles of PBA. The lower mean scores ( $\bar{X} = 4.50$ ) in items 39, 40 and 41 ,when respectively interpreting, revealed that PTs were effective assessment method, even more effective than TA and these were in accordance with the courses taught at that preparatory school. The lowest mean value in item 44\* ( $\bar{X} = 1.50$ ) also stressed that both administrators approved of the consistency between PTs and course content. The second lowest value ( $\bar{X} = 3.50$ ) of this section in the 45<sup>th</sup> item showed that both

participants agreed that PT practices were useful as a supplementary tool to TA.

According to this mean value, they were slightly undecided about the issue.

**Table 31.** Opinions of administrators on “program consistency” phase after the PT implementation

Items	N	$\bar{X}$	SD
39. Performance-based assessment is an effective assessment method for ELT.	2	4,50	,71
40. Performance-based assessment methods are more useful than traditional exams for language learning and teaching.	2	4,50	,71
41. PT implementation applied in this institution was consistent with the content of the courses.	2	4,50	,71
42. Performance-based assessment elements were in accordance with the objectives of the language program of this prep-school.	2	5,00	,00
43. PT implementation in this institution reflected the principles of performance-based assessment.	2	5,00	,00
44. PT implementation conducted in this institution did not reflect what students have learnt.*	2	1,50*	,71
45. Performance-based assessment is useful only as a supplementary tool to traditional assessment.	2	3,50	,71
<b>TOTAL</b>	<b>2</b>	<b>4.50</b>	<b>,51</b>

\*Reverse item

**4.3.**What were the views on the continuity of the current performance assessment implementation as stated by

a. students?

b. instructors?

c. administrators?

In the open-ended section of the Likert-scale 129 students from three language levels (elementary, pre-intermediate and intermediate) were asked to express their opinions about the continuity of the PT implementation in the upcoming years. Three options were given for the respondents. (1) If they chose the first option, they thought that PTs should continue in the same way as in the fall term and would give reasons for that. (2) If they selected the second option, they thought that PT practices should continue with some revisions and they would make suggestions. (3) If they selected

option three, they thought PTs should not continue in the following years and they would state their reasons. Qualitative data gathered from these participants were analysed through content and thematic analysis and the same themes emerged at the end of the analysis, which were: (A) Learning Aspect, (B) Language-related Aspect, (C) Affective Aspect, (D) Assessment-oriented Aspect, (E) Task-specific Aspect, and (F) Application Aspect. In some cases, there were no responses for the themes. The subcategories and corresponding percentages of respondents are presented in tables.

The participant stakeholders' responses for the continuity of the present PT practices revealed that administrators, a vast majority of instructors and a great number of students agreed on the continuity of these practices. In total two administrators, 97% of instructors and 74% of students approved of the PT implementation for the upcoming practices.

#### **4.3.1** What were the views on the continuity of the current performance assessment implementation as stated by students?

In total 22% of participant students believed that PT implementation should continue as the same way as it was administered to them. The total percentages varied by their language levels. 30% of the students from intermediate level, 22% of the students from pre-intermediate level and 17% of the students from elementary level agreed with the PTs' continuity in the same way and gave their reasons.

According to the findings presented in table 32, three main reasons related with learning were given. 3% of students thought that the PT implementation should continue due to its contribution to their reinforcement of what they learnt in the courses. 2% of students believed that PTs should go on in the same way because they trigger students to learn and practice English more. The same number of students

thought that the current PT practices in fall semester encouraged their autonomous learning.

The language-related reasons had three subcategories. 8% of students favoured the current PT implementation because they thought PTs caused an improvement in their general English proficiency. 4% of students mentioned specific language skills and elements that were developed through PTs. 1% advocated the current PT implementation due to the use of more authentic language through PTs.

Three reasons were stated related with the affective factors. 5% of students believed that the current PT practices motivated them to learn English. 2% of students agreed with the continuity of PTs in the same way since they believed that PTs were less stressing than TA practices. Only 1% of students showed self-confidence as a reason for the continuity of the existing PT practices without any changes.

There was only one reason mentioned in terms of assessment. 4% of students favoured the current implementation because of its scoring contribution to their general point average (GPA).

In terms of application, two reasons were stated. 7% of students thought that the current PTs had a positive impact as a useful application for students so they should go on like that. 2% of participants claimed that PTs should continue in the same way since it was an easy application to follow and it was rather systematic.

No task-specific reasons were mentioned by the respondents.



**Table 32.** Reasons of students for the continuity of PTs in the following years in the same way (n=129)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>		
Reinforces course content	4	3%
Triggers students to learn and to produce	3	2%
Renders students' autonomous learning	3	2%
<b>B. LANGUAGE-RELATED ASPECT</b>		
Improves general English proficiency	10	8%
Improves specific language skills & areas	5	4%
Encourages more authentic language use	1	1%
<b>C. AFFECTIVE ASPECT</b>		
Increases students' motivation	6	5%
Lower stress level for students (than TA)	3	2%
Increases students' self-confidence	1	1%
<b>D. ASSESSMENT ASPECT</b>		
Contributes to students' GPA	3	4%
<b>E. TASK-SPECIFIC ASPECT</b>		
	0	0%
<b>F. APPLICATION ASPECT</b>		
PT is a useful application	9	7%
Easy to follow & systematic application for students	3	2%

In total 52% of students believed that the current PT practices should continue with some revisions and they made some suggestions to improve these practices. In terms of language levels, students' percentages of this agreement ranged. 54% of students from pre-intermediate level, 50% of intermediate level students and 52% of students from elementary level stated that they favoured PTs but they need some alterations.

As presented in Table 33, students made three learning-related suggestions. 2% of students favoured PTs but they thought PT implementation should be assisted with more in-class activities and tasks. The same percentage of students suggested that PTs should be more informative. 1% of students thought that the current PT implementation was not useful enough to develop their personal skills and this should be improved.

The participant students made four main recommendations related with language. 7% of students thought that PTs administered for reading course should be improved. 3% of students proposed that the focus of PT practices should be on language rather than craftiness. 2% of students believed that vocabulary PTs should be improved for better practices. Only 1% of students suggested that students should produce more through productive PTs.

In terms of affective issues, only one suggestion was stated by the respondents. 1% of students thought that they needed some training to increase their self-confidence and motivation.

The respondents produced two recommendations related with assessment. 5% of students suggested revising the scoring percentage of PTs. They mainly thought they deserved more points corresponding to their effort. 3% of students claimed that PTs should be on voluntary basis and should not be assessed.

Students' suggestions in terms of task-specific issues formed six subcategories. 13% of students urged that PTs should be designed as more interesting, creative and authentic tasks. 10% of students thought that PTs should not include in-class presentations. The same percentage of students claimed that materials to present PTs such as CDs and posters should be changed. 9% of students suggested lowering the number of PTs currently carried out. 7% of students proposed that PTs should be easier for them while 4% of students thought that they should be provided more flexibility and variety in PT implementation.

Participants suggested six revisions in terms of application. 6% of students thought that they needed more time to complete PTs and it should be changed. 6% of students believed that for each PT there should be time intervals. 4% of students

requested more guidance before and during the PT process while 3% of participants thought that in the following PT practices, the timeline should be determined so as not to clash PTs with exams. 2% of students proposed that PT implementation would be more useful if assessment week should be planned more effectively whereas the same percentage of students seemed to favour PTs but they thought PT application process took a long time so it should be revised accordingly.

**Table 33.** Suggestions of students for PT implementation in the future (n=129)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>		
There should be more in-class tasks & activities to support PTs	3	2%
PTs should be more didactic	3	2%
PTs should encourage more personal skills development	1	1%
<b>B. LANGUAGE-RELATED ASPECT</b>		
PTs for reading skill should be revised	9	7%
PTs' focus should be on English language (instead of craftiness)	4	3%
PTs for vocabulary course should be revised	3	2%
PTs should serve for more language production	1	1%
<b>C. AFFECTIVE ASPECT</b>		
Students should be trained to increase their self-confidence & motivation	1	1%
<b>D. ASSESSMENT ASPECT</b>		
The scoring percentage should be revised	6	5%
PTs should not be scored	4	3%
<b>E. TASK-SPECIFIC ASPECT</b>		
PTs should be more interesting & creative & fun & authentic	17	13%
PTs should not be presented in front of the class	13	10%
Materials (CDs & posters) to present PTs should be changed	13	10%
The number of PTs should be less	12	9%
PTs should be less difficult & less demanding	9	7%
More variety & flexibility should be given to students in PT practices	5	4%
<b>F. APPLICATION ASPECT</b>		
More time should be allowed for students to complete PTs	8	6%
Separate deadlines should be determined for each PT	8	6%
More guidance should be provided for students before & during the process	5	4%
PTs should not clash with the exams	4	3%
Assessment week should be planned in a more effective way	3	2%
PT application should be less time-consuming for students	3	2%

In total 26% of students disapproved of the current PT implementation and gave their reasons to support the idea. The percentages of students ranged from language level to language level. 35% of students from elementary level, 24% of pre-

intermediate level students and 20% of students from the highest level at the preparatory school rejected the idea of PTs' continuity for some reasons.

As seen in table 34, students stated only one reason by a student (1%) related with learning in the negative direction of PTs' continuity in the following years because of its cognitive level which is claimed to be unsuitable for students.

Four reasons were stated by the participants in terms of language. 9% of students did not favour PTs since they thought PTs did not contribute to their language learning. 3% of participants believed that PTs should not continue because they did not encourage language productivity. 2% respondents expressed negative opinion on PTs due to students' incapability of language production. 1% of students complained that PTs' focus point was not on language production.

Two reasons were expressed by students for not favouring the current PT implementation in terms of affective issues. 5% of students suggested that they did not have enough motivation to prepare PTs. 4% of students claimed that PTs should not be applied in the following years since they caused a lot of stress for students.

Students also gave three assessment-oriented reasons to object to the continuity of the current PT implementation. 5% of students showed a TA tendency as a reason to disapprove the implementation. 3% of students had some scoring concerns so they did not wish PTs to continue. 2% of students were worried about PTs' contribution to their GPA. They probably thought that they might lose scores because of their failure in PTs.

The task-specific reasons for students' disapproval consisted of three subcategories. 3% of students thought that video-shooting requirement of PTs was

useless. 2% of students believed that PTs covered only limited content so they did not like them. The same percentage of students did not want PTs to continue in the following years because while focusing on PTs, they were distracted from their courses and exams.

Students' reasons related with application had two subcategories. 23% of students complained about the time and effort required for this application. 3% of students were quite unwilling to participate in these practices because PTs were obligatory. As PTs were scored, they would lose points if they failed.

**Table 34.** Reasons of students for not continuing the PT implementation (n=129)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>		
Not suitable for students' cognitive level	1	1%
<b>B. LANGUAGE-RELATED ASPECT</b>		
No contribution to students' language learning	12	9%
Encourages rote learning rather than productivity	4	3%
Students' incapability of producing language	3	2%
Focus should be on English language rather than craftiness	1	1%
<b>C. AFFECTIVE ASPECT</b>		
Lack of motivation for students to prepare PTs	6	5%
Higher Stress level for students	5	4%
<b>D. ASSESSMENT ASPECT</b>		
TA preference of students	6	5%
Students' concerns about PTs' scoring	4	3%
Students' concerns about PTs' contribution to their GPA	2	2%
<b>E. TASK-SPECIFIC ASPECT</b>		
Video shooting is a waste of time	4	3%
Too limited to certain topics & content	3	2%
Distraction from courses/exams	3	2%
<b>F. APPLICATION ASPECT</b>		
Requires extra time & effort for students	30	23%
Resistance/Unwillingness to compulsory application	4	3%

**4.3.2.** What were the views on the continuity of the current performance assessment implementation as stated by instructors?

In total 12% of instructors believed in the continuity of PTs in the following years in the same way while 85% of instructors urged some revisions for the

continuity. 3% of instructors objected to the upcoming PT practices by giving their reasons.

Among all participant instructors, 12% of them favoured the current PT implementation and they claimed that it should continue without any changes.

In terms of the learning seen in table 35, instructors stated five reasons for their approval. 12% of instructors supported the present PT implementation because it encouraged students' autonomous learning. 10% of instructors thought that PTs should go on since they provided a deeper insight of learner progress for both instructors and students themselves. 2% of instructors advocated PTs due to its contribution to experiential learning of students. 2% of instructors thought that PTs gave opportunity for students for their reinforcement of course content while the same percentage of instructors believed that PTs developed students' personal skills.

According to instructors' responses in terms of language, they mentioned three reasons to advocate the continuity of PT practices in the same way. 5% of instructors thought that through PTs students' integrated language skills improved. 5% of instructors believed that PTs developed students' general English proficiency while 3% of them thought that PT practices were useful for fluency.

Two affective reasons were given by the participants. 3% of instructors considered PTs as a fun application for students whereas the same percentage of them thought that TA caused more stress than PTs for students. Therefore, they advocated the continuity of these practices without any alterations.

Only one task-specific reason was mentioned. 2% of students advocated the current PT implementation its creativity factor for students.

Instructors did not mention any reasons related with assessment and application for PTs' continuity in the following years in the preparatory school.

**Table 35.** Reasons of instructors for the continuity of PTs in the following years in the same way (n=60)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>		
Renders students' autonomous learning	7	12%
Provides a deeper insight for students' learning process	6	10%
Encourages experiential learning	1	2%
Reinforces course content	1	2%
Students' personal skills development	1	2%
<b>B. LANGUAGE-RELATED ASPECT</b>		
Improves integrated language skills	3	5%
Improves general English proficiency	3	5%
Effective way to assist fluency	2	3%
<b>C. AFFECTIVE ASPECT</b>		
Enjoyable application for students	2	3%
Lower stress level for students (than TA)	2	3%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>		
<b>E. TASK-SPECIFIC ASPECT</b>		
Increases students' creativity	1	2%
<b>F. APPLICATION ASPECT</b>		

In total 85% of participant instructors agreed that the present PT practices should continue with some revisions. As seen in table 36, their suggestions to improve the current implementation accumulated in only three themes. Instructors mentioned no suggestions in terms of learning, language, and affective issues.

Eight main assessment-oriented suggestions were made by instructors with low percentages compared to their application and task-specific suggestions. The highest percentage (17%) in this section belonged to instructors' suggestions on scoring rubrics. They thought rubrics required some alterations. 7% of instructors suggested that assessment of PTs should be more objective. 7% of these claimed that PTs required assessment training. While 5% of participants suggested scoring in-class

PT presentations, the same percentage of them suggested using one rubric for all PTs. %5 of them thought that students did not present their original work and it should be avoided. 3% of instructors proposed that grammar might be measured through written output rather than spoken. %2 of them suggested peer-evaluation for PTs.

Instructors produced seven Task-specific suggestions. The majority of the overall section by 35% urged that the number of PTs carried out should be lessened. 13% of instructors recommended that it would be better if some PTs were merged. The same number of participants thought that preparing video-shooting for PTs should be changed. 10% of them claimed that tasks should be revised and created in a more efficient, authentic and creative way. 8% of instructors thought that the number of task options should be changed. They claimed that some PTs had a lot of options while others did not have sufficient options. 7% of them believed that PTs were difficult and time-consuming so there should be a revision in terms of practicality. 5% of instructors wanted PTs to include more pair and group-work tasks.

Instructors stated fourteen suggestions related with application. 20% of instructors suggested facilitating technology more during the PT practice whereas 15% of them suggested PTs not to be an obligatory application. 8% of them thought that it was not effective for students to hand in their PTs at the same deadline. 7% of participants requested providing more guidance to students during PT process. 5% of them thought more time should be given for instructors to assess PTs. 3% of instructors thought that both the product and preparation should be scored. 3% of respondents respectively suggested lowering class population for an application like that, allowing more time for in-class PT presentations, PT completion and PT explanations for students. The same number of instructors (3%) suggested that students should not be distracted from their lessons due to PTs. Therefore, they should



not take long time to prepare these. %2 of participants respectively suggested there should be more frequent PT practices, instructors should have more control on these practices and good samples of PTs should be displayed to students to increase motivation and interest.

**Table 36.** Suggestions of instructors for the PT implementation in the future (n=60)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>	0	0%
<b>B. LANGUAGE-RELATED ASPECT</b>	0	0%
<b>C. AFFECTIVE ASPECT</b>	0	0%
<b>D. ASSESSMENT ASPECT</b>		
Changes are needed in the scoring rubrics	10	17%
Subjective assessment should be avoided	4	7%
Assessment training should be provided for instructors and students	4	7%
Presentations of PTs should be scored	3	5%
One rubric for all tasks should be designed	3	5%
Plagiarism should be avoided	3	5%
Grammar should be assessed through written output	2	3%
Peer-evaluation should be integrated	1	2%
<b>E. TASK-SPECIFIC ASPECT</b>		
Number of PTs should be less (Not for all skills separately)	21	35%
Some PTs should be integrated	8	13%
PTs should be prepared rather than video shooting	8	13%
More creative & efficient & authentic tasks should be designed	6	10%
The number of task options should be changed	5	8%
Less difficult & time-consuming tasks should be designed	4	7%
PTs should include more pair or group-work	3	5%
<b>F. APPLICATION ASPECT</b>		
Technology should be used more for collecting & assessing PTs	12	20%
PTs should be an optional application	9	15%
PT deadlines should not be determined at the same week	5	8%
More guidance should be provided for students	4	7%
More time should be allotted to scoring	3	5%
Class population should be lower for PT application	2	3%
Not only product but also the preparation process should be scored	2	3%
More time should be allotted to presentations	2	3%
More time should be allotted to students to complete PTs	2	3%
More time should be allowed to explain tasks to students	2	3%
Distraction from lessons should be avoided	2	3%
PT application should be more frequent	1	2%
Instructors should have more control on PT application	1	2%
Good PTs should be displayed as samples	1	2%

In total 3% of instructors disapproved the continuity of the present PT implementation by submitting some reasons. No instructors mentioned learning aspect of the practices negatively.

As seen table 37, instructors had two language-related reasons. While %2 of instructors commented that PTs did not contribute to students' language development, the same number of the respondents claimed that students' PTs were disappointing.

Only one affective reason was stated for the objection. 2% of instructors thought that PT implementation should not continue since students did not get motivated and did not pay attention to these practices.

Participants gave three assessment-oriented reasons. %2 of instructors respectively expressed negative opinions on PTs because they thought that PTs were not easy to assess for them, assessing PTs took a long time and these tasks were difficult to score them objectively.

Task-specific reasons of instructors formed two subcategories. %2 of instructors believed that PT practices should be eliminated since their number was too demanding. %2 of them thought that they caused paper waste.

In terms of reasons related with application, instructors stated one reason to show their disapproval. 3% of instructors did not favour PT practices because they claimed that these tasks were not administered as process-oriented practices as they were planned before the actual application.

**Table 37.** Reasons of instructors for not continuing the PT implementation (n=60)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>	0	0%
<b>B. LANGUAGE-RELATED ASPECT</b>		
No contribution to their English language development	1	2%
Disappointment by the poor language quality of PTs	1	2%
<b>C. AFFECTIVE ASPECT</b>		
Lack of motivation & attention of students for PTs	1	2%
<b>D. ASSESSMENT-ORIENTED ASPECT</b>		
Difficult to assess	1	2%
Takes too much time to assess	1	2%
Difficult to assess objectively	1	2%
<b>E. TASK-SPECIFIC ASPECT</b>		
Overwhelming number of PTs (not for all courses)	1	2%
Too much paper waste	1	2%
<b>F. APPLICATION ASPECT</b>		
Not a process-oriented application as planned	2	3%

#### 4.3.3. What were the views on the continuity of the current performance assessment implementation as stated by administrators?

As both administrators selected the continuity of PTs with revisions option and commented on it, only these suggestions were presented in Table 38. According to that option they advocated that the current PT implementation should continue for the following years at the preparatory school but some revisions were needed.

Participants had no suggestions to improve the current PT implementation in terms of learning, language, affective and assessment-oriented issues.

As presented in Table 38, administrators made three task-specific suggestions. Both administrators proposed that PTs should be revised in accordance with the new developments in the field. One of them suggested that the variety of PTs should be improved whereas one of them suggested increasing the aimed effect of this application by the help of these constant revisions.

Administrators made one suggestion about application. One of the executive staff members highlighted that collecting feedback from stakeholders constantly should be effective to enhance the present PT implementation.

**Table 38.** Suggestions of administrators for the PT implementation in the future (n=2)

THEMES & SUBCATEGORIES	f	Percentages of res.
<b>A. LEARNING ASPECT</b>	0	0%
<b>B. LANGUAGE-RELATED ASPECT</b>	0	0%
<b>C. AFFECTIVE ASPECT</b>	0	0%
<b>D. ASSESSMENT ASPECT</b>	0	0%
<b>E. TASK-SPECIFIC ASPECT</b>		
PTs should be updated & revised constantly	2	100%
The variety of PTs should be increased	1	50%
The aimed impact of PTs should be increased	1	50%
<b>F. APPLICATION ASPECT</b>		
Feedback should be collected from instructors & students constantly	1	50%

#### 4.4 Summary of Findings

According to the opinions of the participant stakeholders in terms of positive and negative aspects of performance assessment implementation, very similar ideas and themes were proposed. All stakeholders densely mentioned as their positive expectations related with learning and language. The most prominent expectations were benefits of PTs to learner autonomy, higher-order skills, and different ways of learning and retention of the courses. While instructors had more positive expectations about assessment-oriented issues, students mostly had positive expectations about language development, authentic language use and improvement in specific language skills. In terms of assessment, PTs were considered as a more flexible practice and a contribution to GPA by the students while instructors wished to facilitate PT practices to monitor students' progress. Administrators highlighted the process-oriented aspect of performance assessment. They also expected PTs to be less stressful, particularly compared to TA practices, more enjoyable and a trigger for

students' self-confidence. Participants did not mention any positive expectations related with tasks and their application.

From the negative perspective, none of the participants stated any negative anticipation of PTs in terms of language and learning, indicating that they all believed in their contribution to these areas before the implementation. However, both instructors and students predominantly highlighted assessment-oriented concerns, specifically students were concerned about the time and effort required for PBA while instructors' responses displayed a high level of concern in reliability, validity and subjectivity. The existence of motivation and joy-related ideas seen both in the positive and negative comments by instructors and students revealed that the stakeholders were not sure how this process would feel before the implementation. After assessment issues, concerns gathered around task-specific and application-related factors as the specific details were unknown to the participants. Both instructors and students predicted a need for training and guidance before and during the PT process whereas administrators mentioned minor concerns such as possible technical problems and paper waste related with the tasks. Administrators predicted that PT implementation could be a workload requiring time and effort for the instructors as the instructors themselves seemed to be aware of it.

According to the findings of the scales, students' subscale mean values ranged between 4,12-3,35, instructors' subscale mean values range between 4,18-3,61 and administrators' subscale mean values range between 4,87-4,50. Findings related with the stakeholders' evaluation of the overall PT process in five phases after the actual practice indicated that students liked planning and application phases while instructors and administrators liked planning and scoring phases of the PT practices. While students hesitated about the effectiveness of PTs in terms of learning outcomes and

their consistency with the language program, instructors also favoured the learning outcomes phase minimally. In terms of specific skills, all stakeholders believed in PTs' positive contribution to reading skills, speaking skills and lexical knowledge of students whereas instructors and students both agreed that PTs did not contribute to the development of students' listening skills and structural knowledge. In terms of learner autonomy, reinforcement and self-confidence, instructors mostly expressed more positive opinions whereas students were indecisive about PTs' positive impact particularly on motivation and self-confidence. While students were quite positive on PTs' consistency with course content and program objectives, instructors had more positive opinions of the overall program consistency. The findings revealed that both instructors and students needed more time to compare TA and PBA effectively. The effectiveness of PBA was sensed in the responses of all stakeholders. Although administrators, who had some doubts of the guidance provided to students during the PT process, expressed generally strong positive opinions in all five phases of the implementation, relatively lower scores by students and instructors displayed that they had some doubts on the effectiveness of PBA as a supplementary tool.

The participant stakeholders' opinions on the continuity of PT implementation showed that all administrators, almost all instructors and a great number of students expressed their positive opinions in the direction of continuity. In total two administrators, %97 of instructors and 74% of students said yes to PTs' future implementation. A little more than half of the students (52%), a great number of instructors (85%) and both administrators suggested various ideas to revise and improve the implementation. These recommendations seemed quite detailed and information-loaded. The instructor participants who voted for no alteration mainly favoured language development and learning contribution of PTs; additionally,

students highlighted motivation and scoring advantage. Suggestions from both instructors and students densely accumulated in the task-specific and application aspects of the PT practices, with the instructors providing assessment-related suggestions as well. The suggestions on improvement after the actual PT practice also revealed that the participants had more sophisticated and various thoughts about PBA. Instructors mostly urged revisions in PT scoring rubrics while students suggested revisions on reading and vocabulary course tasks. The suggestions about tasks were mainly related with the number, difficulty and content. Both students and instructors highlighted the need for more time and guidance for PTs. More technology use was the common point of both participants as well. Participants' emphasis on time and effort required for PTs shifted from assessment-orientation to application since they seemed that time and effort issues were directly related with this specific application of PBA rather than the assessment method itself. The participants who disapproved the continuity of PTs (students 26%, instructors 3%) had common reasons such as stress, lack of motivation, too much time and effort. Additionally, as assessors, instructors mentioned the scoring difficulty and subjective scoring as the reasons of their disapproval. Administrators, who seemed quite satisfied with the current PT implementation according to the scale results, suggested revising PTs and increasing their variety. They also stressed the necessity of constant feedback collection from stakeholders for the upcoming practices of this assessment method.

## CHAPTER 5 DISCUSSION

In this section, the whole PT implementation process will be evaluated based on the decision/accountability-oriented evaluation approach by Stufflebeam (2001) in the order of its relevant questions. In addition, the findings will be discussed and interpreted in the light of previous research. The conclusion part includes the presentation of possible impact of the study on language education and some recommendations for future research in the field.

### 5.1. Evaluation of the Evaluation

Performance-based assessment component of the language program under the investigation of this study was evaluated in the light of decision/accountability-oriented approach (Stufflebeam, 2001). The main purpose of this formative evaluation was to evaluate and improve the assessment component and thus the current language program in respect of participant stakeholders' opinions and suggestions on PT practices. Relevant questions are to be responded during evaluation process (Weiss, 1998). Therefore, the questions addressed by this approach (Stufflebeam, 2001, p. 56-57) are answered accordingly.

*Has an appropriate beneficiary population been determined?*

At the onset of the study, the main beneficiary population was determined as adult EFL learners at a preparatory school of a state university in Turkey. In addition, instructors and administrators were seen as a part of the beneficiary population as well.

*What beneficiary needs should be addressed?*

The beneficiary population needed more oral and written production in the target language. In other words, particularly productive skills of students were aimed



to improve through PTs. The other aim was to increase students' autonomy and motivation. More learner-centered setting was needed in preparatory schools in Turkey (Sönmez, 2013; Özeren, 2013). The learner performance was to be assessed by real performance assessment practices.

*What are the available alternative ways to address these needs, and what are their comparative merits and costs?*

The alternative ways to address these learner needs were the improvement of TA practices currently used in the context or the use of different kinds of alternative assessment such as authentic assessment or portfolio assessment. As portfolio assessment, which was not in the scope of this research, was carried out in the writing course of that institution, performance-based assessment was determined as an alternative assessment method for the other courses. As mentioned in the methodology section, the high number of voluntary participation to these performance tasks in the trial year was a promising motive to continue the PT practices in combination with already existing TA methods. There were several comparative merits of PBA as an alternative assessment method such as involving real life and authentic activities (Weigle, 2002). PBA not only encourages authentic language use, but also measures higher-order skills that cannot be assessed via TA (Perlman, 2003) by assisting student autonomy and motivation and above all assessing productive skills of learners which cannot be measured by TA. However, this study advocated the effectiveness of the combination of both assessment methods as supported by some study results in the field (Gökçen, 2015; Göçtü, 2013).

*Are plans of services and participation sound?*

Although the weak points of PBA were taken into consideration, the advantages of PBA and its effectiveness were supported by scientific information and

research as mentioned in the literature review section. The overall PT implementation was designed stage by stage in accordance with the basic phases of program evaluation: 'Planning, application, and evaluation' (Topkaya & Küçük, 2010).

Throughout the evaluation process, sufficient training and even warning notices were provided for the participants in order to prepare sound circumstances for a more valid implementation. The participation of the stakeholders was ensured through the consent forms as university policy and in terms of research ethics.

*Is there adequate provision for facilities, materials, and equipment?*

Since PTs require significant resources (Önal, 2010), necessary resource reference lists were announced to students such as book names or web-sites to facilitate in the PT process. For these resources, students were directed to the self-access center of the institution as well. However, the provision for materials and equipment such as posters and CDs were students' responsibility.

*Is the program staff sufficiently qualified and credible?*

All participant stakeholders had sufficient training in ELT. Additionally, assessment, alternative assessment and specifically PBA training was provided to both practitioners and students before and during the PT process. Instructors were provided with some seminars, one of which was carried out by an expert in assessment, Professor Farhady before scoring PTs since training is essential for PBA. For instance, Linn, Baker & Dunber (1991) criticized PBA for several reasons including its requirement for training. For ongoing improvement, in spring semester in the light of instructors' suggestions, more practical assessment training was planned by Professor Farhady.

*Have appropriate roles been assigned to the different participants?*

Throughout the whole PT implementation process, the roles of the participants were predetermined. The administrators were in charge of directing, organizing, controlling and monitoring the process while the coordinators were responsible for designing performance tasks in pursuance with the course content, language program objectives and student language levels and also preparing the necessary documents such as scoring rubrics and evaluation sheets to conduct the process. All these necessary equipment was checked by the administrators before the actual practices. Although the contact meetings were organized by the administrators, they were held under the chairmanship of the coordinators. The researcher of this study, who was the current assistant director of the institution, did not attend any meetings as a chairman in order to sustain her objectivity and not to cause any biased results. The instructors, who were responsible to carry out the trainings of students, were the practitioners of the PT practices in class. They were supposed to adopt a role of being a guide for the students during the process. They were also in charge of scoring the PTs and monitoring the students' presentations on these practices. Alternative assessment renders students to have an active role in their learning (Çetin, 2011). In accordance with the targeted outcomes of PTs, students were supposed to adopt an active and more autonomous learner role throughout the PT process.

*Are the participants effectively carrying out their assignments?*

All participants, administrators, participant instructors including coordinators in charge carried out their assignments properly throughout the process of the evaluation. The students stated that they did their PTs themselves and could hand them in on time. These statements were also approved by the instructors and

administrators during the evaluation meeting held after the PT implementation and supported with the data findings of this study.

*Is the program working and should it be revised in any way?*

During the implementation period there were frequent checks on the running of the program through various means. In training-oriented contact meetings informal feedback was collected from instructors and through the qualitative and quantitative data collection instruments utilized in the study, sound and scientific data was gathered from all stakeholders to monitor whether the target assessment component of the current language program was working or not. As defined by Orstein and Hukins (1998), program evaluation is a process which requires modifications and alternations. Therefore, it was planned to share the study findings with the stakeholders and open those to discussion for the necessary revisions in order to improve the program (Appendix K).

*Is the program effectively reaching all the targeted beneficiaries?*

The main purpose of applying PTs in this prep-school was to activate what students learnt in their courses through oral or written out-of-class tasks via a performance or a product. Another goal of PTs was to increase learner autonomy and motivation to learn English. Through a more-learner-centered assessment implementation, in order to achieve these goals, all stakeholders should be reached effectively. For the evaluation a big sample of preparatory school students consisting of 129 students were used. In addition, the whole staff as practitioners and the administrators in charge were engaged in the evaluation process.

*Is the program meeting the participants' needs? Did beneficiaries play their part? Is the program better than competing alternatives?*

Throughout the process of PBA implementation and data collection periods, it was observed that the assessment component of the program met the participants' needs as a complementary tool to the current traditional assessment practices. The use of PBA in the present context has gained quite positive interpretations compared to TA only. Findings indicate that students as well as instructors and administrators were pleased with the use of PBA as a supplementary assessment tool. In other words, data showed that instructors and students tend to favor the combination of both assessment methods while administrators seemed to advocate PBA more than TA.

*Is it affordable? Is it sustainable? Is it transportable? Is the program worth the required initial investment?"*

Considering the application phase of the PTs for students, the unaffordable aspects of PBA based on the data collected were supplying the necessary equipment such as CDs and posters for the students and spending too much time and effort for both the instructors and students. In general, alternative assessment methods require great effort to plan and apply and seem quite time-consuming (Zaimoğlu, 2013). Alternative assessment practices are usually too demanding for teachers (Farhady, 2003). However effective practices advocate variety of assessment (Nasab 2015). Therefore, the gains such as learner autonomy, more language production and increase in general language proficiency may compensate for its weak points. Although Stufflebeam stated that providing necessary conditions of PBA is difficult (2001), it may not be hard to provide these conditions if the learner needs, language and cognitive level and also their interests in the target context are carefully taken into consideration. When the gains for the sake of language learning, other learning

outcomes and positive stakeholder opinions are considered, performance-based assessment component of that language program seems to be worth the required initial investment. There were no major problems encountered in this respect in the current study.

## **5.2. Discussion**

The results of this research are discussed in terms of the research questions respectively. The prominent points of these findings are interpreted in conjunction with the relevant research in the literature from a broader aspect.

5.2.1. What were the potential positive and negative aspects of performance assessment tasks before their implementation from the perspective of students, instructors and administrators?

According to students' responses on positive aspect of PTs before the practice, students mostly focused on the contribution of PTs to their learning particularly in terms of their autonomous learning and personal skill development. In accordance with the linguistic development gain through alternative assessment claimed by Kavaliauskienė & Anusienė (2007), almost half of the students believed in PTs' potential positive impact on their general language proficiency. Their general expectation from PTs to increase their self-confidence and motivation seemed quite low. Some students anticipated PTs to be enjoyable and not stressful. These affective reasons might be related to their assessment-oriented responses since nearly one-fifth of the students thought that the flexibility in terms of time would give them the opportunity to check their mistakes and correct them, which is not valid in one-shot pen and paper exams administered in limited time. These findings may correspond to Burnaz's claims (2011) on the time-pressure, demotivation and anxiety caused by the

TA practices. Almost the same number of students in the study considered PTs as a chance to increase their GPA. It is quite natural that students did not have strong specific claims on tasks and their way of application before experiencing them. Considering the negative opinions of students on PTs, it is clearly seen that the majority of students had assessment-oriented concerns about PTs since Turkish students generally adopt a rooted exam culture not a learning culture as categorized by Hamp-Lyons (2007). Almost half of the students were concerned about the extra time while one-fourth of them complained about the effort they would have to spend for PTs. Particularly students' expected time concerns seem quite parallel with the common negative claim that alternative assessment methods are 'time-consuming' (Al-Nauh, Taqui & Abdul, 2014; Zaimoğlu, 2013). AS an assessment method, PBA is considered as a time-consuming assessment practice by some experts (Perlman, 2003; Brown, 2004). As TA and summative assessment practices are more common in preparatory schools (Doğan, 2009), students' unfamiliarity with this kind of assessment may raise their anxiety level since in total one-fifth of the students mentioned their possible high stress level plus lack of motivation.

According to instructors' expectations on the positive aspect of the PT implementation, learning-related issues of PTs far outweighed others. Vast majority of instructors expected PTs to encourage autonomous learning. Through this new implementation, they expected their students to increase their higher-order skills and to reinforce what they learnt in the courses. On the contrary, according to Zaimoğlu (2013) TA is claimed not to construct reinforcement and retention of the students' knowledge gained in courses. Instructors' responses indicate that they are in search for a more learner-centered setting, which is needed generally in preparatory classes (Sönmez, 2013 ; Özeren, 2013). As TA practices lack measuring students' real

performance and productive skills (Brown, 2001-2003), more than half of the instructors believed that PTs might be a remedy to monitor their holistic progress. One-fifth of instructors expected PBA to provide positive washback and impact in pursuance with student needs. It is a promising claim that PBA adopts an increasingly significant role in English language programs and it reflects objectives thanks to its washback effect (Brooks 1999). According to Moss (1992) PBA has more valid interpretations about its interlocutors compared to TA only. Therefore, if these interpretations are perceived correctly, they may lead to sound revisions in language programs. In addition to its failure to assess performance, Doğan (2009) states that TA fails to measure higher-order skills, which were expected to improve through PTs by the participant instructors. It is suggested by Spector-Cohen that PBA activates higher-level thinking, problem-solving higher-order skills (2007). While more than one-third of instructors predicted PTs might increase student motivation, nearly the same number of them thought that PTs might encourage authentic language use of students. As PBA is a new assessment method for the practitioners, they might not have commented on tasks and the application aspect before the actual practice. Like students, instructors also did not mention any negative opinions on PTs' learning and language-related aspects. This reveals the fact that both of them perceived PTs as a useful method for English language and an opportunity in terms of learning. On the other hand, more than half of the instructors had some concerns about the main dimensions of assessment, defined by Taylor (2006), particularly reliability and validity. For instance, Messick (1992) and Moss (1992) criticize PBA for its lack of validity. Associated with these concerns, almost half of the instructors of the present study were worried about the possibility of subjective assessment, which is also suggested by Perlman (2003). According to the study results conducted by Zaimoğlu



(2013) teachers prefer objective techniques in classroom assessment. However, Brooks suggests that PBA has validity (1999). A significant number of instructors thought they would need training both for themselves and their students probably due to unfamiliarity. While one-fifth of instructors had potential concerns about plagiarism, the same number of them complained about the time and effort to score PTs and also about the students' possibility of demotivation. Both students and instructors had a common point about PTs in terms of time and effort.

Both of the administrators who participated in this study had common expectations with students and instructors on PTs' contribution to autonomous learning of students. Both also stated that PTs might motivate students to learn English. The increase in authentic language use and development of higher order skills were also considered as the positive sides of the PTs for them before the implementation. One of the administrators stressed that PBA might have a positive impact as both process and product-oriented assessment method. Bayram (2015) also suggests that alternative assessment is not only product-oriented but also process-oriented with beneficial washback effect in terms of advantages of PTs. In this respect, administrators seemed to have less concerns than the other stakeholders before the implementation. Considering their positive view, they may agree with Alderson & Lancaster, who claim these applications are more easily integrated into classroom (2001). However, both of them seemed worried about the extra workload for the instructors. As Farhady (2003) claimed that they seemed to be aware that alternative assessment applications are too demanding for teachers. One of them mentioned that PTs might cause paper waste and one predicted that instructors might have some technical problems during PT practices. Linn, Baker & Dunber (1991) also criticized PBA in terms of technical concerns among their various other criticisms.

Administrators expressed more positive opinions than the negative ones before the actual practice. As in Gökçen's study (2015), in which administrators encourage the use of project-based assessment as an alternative assessment method, the executive staff in this preparatory school seems to support new assessment methods based on these preliminary findings.

5.2.2. How were the planning, application, scoring, learning outcomes and program consistency phases of performance assessment tasks evaluated after their implementation by students, instructors and administrators?

From a holistic perspective in terms of the evaluation phases, the study results indicate that students were satisfied with the planning, application and scoring phases of PTs while they did not have strong claims on the learning outcomes of PTs and its consistency with the current language program. These findings overlap with Hacısalıhoğlu's (2013) study results which indicate students are more aware of the application phase than learning outcomes. This may be resulting from their lack of time and experience with PBA. As this study is based on a formative evaluation, students may need more time to observe its contribution to their learning outcomes in the long run. They seemed quite satisfied with the planning phase whereas in terms of application phase they had some doubts on the joy they got while preparing PTs. Their positive attitude in the scoring section indicates that their concerns about the scoring before the actual practice might decrease relatively. According to their positive responses, students were not negative about time to complete PTs, their scoring and complexity level contrary to the results of the research administered by Chen (2007). Generally, students thought that PTs contributed to their lexical knowledge, reading and speaking skills. Contrary to PTs' positive effect on productive skills, they slightly agreed with their contribution to the writing skills. As

portfolio assessment, which was not in the scope of this study, was carried out in writing courses rather than PTs, this may direct students to disregard PTs' effect on their writing skills. Most of them seemed slightly positive about PTs' contribution to their general language proficiency. In consistency with the results of Chen's (2007) study and Ak's research (2013) in which students showed positive responses in terms of language development, participant students believed in PTs' positive impact to their language proficiency. Students seemed neutral about PTs' contribution to their autonomy and reinforcement whereas they were not quite positive in terms of motivation and self-confidence. This research presents a moderate degree of consistency with the results found out in Chens's (2007) and Yıldırım & Örsdemir's (2013) in terms of motivation increased through PTs as students in this study were indecisive about motivation factor. They mostly expressed positive opinions on PTs' consistency with course content and the principles of PBA. However, contrary to students' general preference claim on alternative assessment formats which reduces stress and anxiety (Doğan 2010 & Büyükkarcı 2010), they hesitated to express positive opinions on PBA's effectiveness, particularly compared to TA practices. On the other hand, they were indecisive to consider PBA as only a supplementary tool. These findings yet again may reveal that more time and experience are required for the learners to analyze this new assessment method as Kızılkaya (2014) and Gökçen (2005) also claim that students need more experience and time to have stronger opinions about alternative assessment practices. They may show TA tendency due to their unfamiliarity to this new implementation.

From the instructors' point of view, all these PT implementation phases were more favored compared to the students, with a slight agreement on learning outcomes phase. Particularly their satisfaction in scoring phase is quite promising compared to

their concerns in the preliminary data. Contrary to students, they slightly agreed that PTs were enjoyable for the students. They seemed indecisive about whether it was a time-consuming practice for them or not. Like students, they may need more time and experience to decide on it. Apart from the listening skills and grammar, they had more positive opinions on PTs' positive impact on all language skills and areas. They also had parallel ideas with students in terms of PTs' contribution to students' general English proficiency. While they had some doubts of PTs' effect on student motivation, they slightly favored PTs in terms of their contribution to reinforcement of course content, self-confidence, and student autonomy. Stipek also (2002) suggests that assessing student performance increases their self-confidence. Even though instructors had some doubts about the superiority of PBA over TA, they mostly believed in its effectiveness as an assessment method. Therefore, they may be advocating the combination of both methods. According to the similar findings in a study conducted by Gökçen (2015), teachers seem to be in favor of project-based assessment, a kind of alternative assessment, and consider it as a supplement to TA. Nasab (2015) also believes that effective practices are assisted by the means of variety in assessment. Combination of TA and alternative assessment is also advocated by Göçtü based on his findings in 2013. Instructors' positive attitude towards the statements in the program consistency subscale indicates that instructors in this study thought that this PT implementation was generally consistent with the current language program and its objectives.

As the data collected from students is important for evaluation (Gaies, 1992), data obtained from administrators and practitioners are useful to enhance the program as in a study conducted by Yang (2007). The study findings showed that administrators are the most positive stakeholders about the PT practices carried out in

that institution. All the phases of the PT implementation were strongly favored by these two participants. Although Zaimoğlu (2013) claims that alternative assessment methods require great effort to plan and apply, in this study these two phases were appreciated by the stakeholders, including administrators. Contrary to the other stakeholders, they expressed highly positive opinions on PTs contribution to students' motivation and self-confidence, similar to their expectation in the preliminary data. When probing their moderate opinions specifically, it is found out that they slightly agreed with the guidance provided for students during the PT process and they slightly agreed that PBA is useful only as a supplementary tool to TA. Considering their strong appreciation of this new assessment practice compared to TA, they may think that TA and PBA can be used in a complementary way effectively.

#### 5.2.3. What were the stakeholders' views on the continuity of the current performance assessment implementation?

Although PBA has been a new implementation lasting for 15 weeks in the context, one-fifth of students seemed satisfied with it. Therefore, they wanted the PT implementation to continue in the same way. PTs' contribution to their general language development, specific language skills, motivation, GPA and reinforcement were their most prominent reasons. Considering the rest of the students, only this limited number of students' opinions seemed to be consistent with the positive results in studies conducted by Chen (2007) and by Kadidja (2015) in terms of increased motivation. The same students in this study found the current PT implementation rather effective. On the other hand, more than half of the students approved of its continuity yet with some revisions. Apart from some learning and language-oriented suggestions with low percentages, the majority of students suggested some revisions on the tasks themselves and their way of application. The most prominent suggestions

were about the number, the content, and the deadlines of PTs. The time allotted to PTs, their presentation in class and materials to prepare PTs were also offered to be revised. These results are parallel with negative comments of students on PTs in terms of time to complete PTs, their scoring and complexity level in Chen's study (2007). However, compared to the preliminary findings, their concerns about scoring seemed to decrease according to the scale results. Numbers of tasks and task complexity were also criticized according to the findings of Ak's study (2013). Findings indicate that one-fourth of students disapproved of the continuity of the PT implementation for various reasons ranging from its taking too much time to requiring too much effort. Some students believed that they had no gain in terms of language learning through PTs. However, in total majority of students believed in its effectiveness and its continuity in the upcoming years. As students' sincerity in assessing exams is proved by the study results (Dittmann-Domenichini ; Halbherr & Schlienger, 2014), these findings are quite positive in terms of performance-based assessment practices from students' perspective.

One-fifth of instructors favored the PT implementation in the fall term and wanted it to continue in the same way. The same as the students, instructors mentioned PTs' positive impact on language development and specific language skill improvement as their reasons. Different from students, they highlighted that PTs improved students' autonomous learning and provided a deeper insight for students' learning process. A great majority of instructors agreed that PT practices should go on with some revisions. They mostly suggested lessening the number of PTs, integrating some PTs, revising scoring rubrics and eliminating video-shooting requirement. They also proposed to use more technology, to make PTs optional and design more authentic, creative and less demanding tasks. Creating effective tasks require some

tedious stages (Cohen, 1995). Thanks to these suggestions from instructors, they may be improved in time. Only two instructors disapproved of the continuity of PT implementation for some reasons. According to Yüce (2015), teachers are used to TA and feel comfortable. These instructors may feel uncomfortable with a new assessment practice. Contrary to their colleagues' opinions, they claimed that PTs had no contribution to students' language development and they were disappointed with the poor quality of PTs. They also complained about the overwhelming number of PTs, assessment difficulty, assessment time and paper waste. As the vast majority of instructors believed that PTs should continue with or without revisions, the results and suggestions seem quite promising for the further PT practices from practitioners' perspective. However, in a longitudinal study on EFL teachers in preparatory school they showed quite negative attitude towards PBA after 6 year-implementation due to tasks, rater training and rating scales (Chinda, 2014). Therefore, these suggestions from participant practitioners should be taken into consideration to make necessary revisions constantly.

Administrators, who were quite positive according to both quantitative and qualitative results in the study, supported PT implementation's future practices with some general revisions. They suggested that PTs should be revised constantly and the variety of PTs should be increased. They also believed in the impact of feedback collection from stakeholders throughout the process. As Çeliker (2015) claims decision making is provided through feedback. They stated that the aimed effect of PTs should be increased. The targeted goals of PTs might be shaped and improved with the help of these feedbacks.

### 5.3. Conclusion

As explored in this study, the use of performance-based assessment tasks as a supplementary tool to traditional assessment methods carried out in the target institution achieved its goal, particularly in terms of language development, learning, and consistency with the program in respect to the opinions of the stakeholders, namely students, instructors and administrators.

Before the actual implementation, PTs were perceived as a means of increasing learner autonomy, motivation, self-confidence and reinforcement. However, as the final results were interpreted, the motivation aspect of performance-based assessment seemed not to be very effective contrary to the positive study findings (Kadidja 2015; Yıldırım & Örsdemir, 2013) in the field whereas both instructors and students in this study seemed indecisive about PTs' contribution to learner autonomy. In terms of self-confidence and reinforcement, instructors showed stronger positive opinions compared to students. The PT application was generally considered as an assessment method requiring time and effort from both practitioners' and learners' perspectives. The study findings also displayed that as all stakeholders came from a traditional assessment background, they needed time to adapt to this recent application. It was also obvious that training and guidance were significant components of an effective PT implementation.

Despite these demanding conditions of PBA, it was observed that if necessary guidance and training were provided, the planning, application and also scoring, which is the most worrying part for instructors, could be effectively handled. As supported with the results of this study, the combination of TA and PBA, supported by Gökçen (2015) and Göçtü (2013), may provide more useful gains for language learning in spite of the advocates of TA only in the literature.



The information-loaded and detailed suggestions especially from the instructors and students in this study indicated that task difficulty, task numbers and task contents should be taken into consideration meticulously. Moreover, as assessors, most instructors criticized scoring rubrics. As suggested in the literature, creating effective scoring rubrics and performance tasks require tedious effort. The expert in charge of assessment training also criticized the scoring rubrics and tasks. He suggested developing rubrics based on more scientifically valid criteria such as Common European Framework and he urged to convert the present performance tasks into more authentic ones. Even though a lot of revisions are needed to enhance the current PT implementation, still it was quite promising that majority of the participants advocated that PT practices should continue.

According to Norris (2009) in this era there have been solid demands for accountability testing, outcomes assessment and quality control so language educators are developing a high awareness of program evaluation and its pioneer role in determining how language teaching and learning occurs. This decision/accountability-oriented formative evaluation revealed that although long-term results should be collected, even a fifteen-week small scale implementation may present valuable information about an alternative assessment component and its consistency with its language program. Decision making of the sustainability of any new practice should be supported with the feedback from the interlocutors, as this decision-making and feedback connection was also stressed by Çeliker (2015). Necessary feedback should be gathered from all stakeholders to gain a deeper sight in the practices, to enhance the weak points and to make sound judgements. Similar to the positive study results conducted by Gunn (1999), improvement-oriented evaluation approach may provide a basis for further large-scale studies in the field.

#### **5.4. Implications**

This study's results indicated that the PT implementation moderately achieved its goal in consistency with the relevant language program despite some weak points detected in terms of the task content, task numbers and targeted learning outcomes. The main purpose of all language assessment types is to improve learners' language achievement. The results before the PT implementation and after the actual practice revealed the fact that all participant stakeholders believed in the positive impact of PTs in language learning. They moderately agreed on the effectiveness of the combination of PBA and TA assessment practices.

The first implication of this research can serve for the program developers and evaluators. As in this case, one component of the language program or the program as a whole should be developed and evaluated with the involvement of all relevant stakeholders. The main focus point of this study was improvement in the assessment component of the program in accordance with its objectives and structure. In order to achieve any development, the scientific data should be collected from the participants. Any new application requires a trial before the actual practice. The impact of a novice application should be investigated scientifically to put on-going revisions into practice. All study results should be shared with the participants to ponder upon and make solutions collaboratively. Although some experts advocate external evaluation in the control of an outsider, an internal researcher who has a solid grasp of the context might also carry out a program evaluation. In this respect, the evaluation model of this study can be used as a sample by program evaluators.

The second implication of this study is for the administrators, who are in charge of integrating new practices into the current language program. First step is to provide necessary training for all participants exposed to the new practice.

Considering the appreciation experienced in this study in terms of the planning and application phases of the PT practices, the administrators should plan and monitor the whole process constantly and ensure that all stakeholders are informed. For instance, the idea of designing a weekly schedule was found quite effective for the participants to follow the process. As suggested in this research, contact meetings before the practice and post-practice meetings should be organized for the staff to detect the weak points of the new implementation and to revise these points in cooperation.

The third category of implications is addressed to the teachers who are willing to design performance-based assessment tasks and its related documents such as scoring rubric, tracking forms and evaluation sheets. Considering the task-specific problems detected in this research, performance tasks should be designed as more authentic and creative tasks to motivate the target learners. They should constantly be revised in accordance with the feedback and suggestions collected from both students and practitioners. The instructions for each PT should be given clearly enough for all participants to follow the process easily. The scoring rubrics should be based on theoretical background and sound criteria.

The fourth type of implication could be valid for the actual practitioners of this kind of assessment application. A new assessment method like PBA, with which most of the students belonging to an exam culture are not quite familiar, can cause some concerns particularly before the actual practice. Therefore, the practitioners should follow the process very carefully and should guide the students before and during the whole process. In addition to the pre-process and while-process assistance, students should be provided necessary feedback for their PTs after their presentation.

The last group of implications is for the language teaching institutions ranging from primary education to higher education which are planning to apply more

learning-oriented assessment methods in order to improve their students' productive skills and autonomy as active learners. The institutions adopting traditional assessment methods should seek various ways of alternative assessment and should have a tendency for a combination of traditional and alternative assessment practices in order to facilitate from the advantages of both methods. Providing necessary conditions for any kind of alternative assessment is not very easy. Therefore, the results of this research can guide these intuitions in terms of planning, application, scoring and program consistency phases of the PT implementation. The weak points of PTs and their targeted learning outcomes should be taken into consideration and should be improved in accordance with the needs and objectives of the relevant language program carried out in these institutions. The institutions should seek not only contemporary assessment methods but also any developments in ELT to serve for a more learner-centered setting.

### **5.5. Suggestions for further studies**

In addition to these implications abovementioned, some suggestions should be made for future studies. As found out in this study, students might require more time and experience to evaluate the PT process, particularly the learning outcomes whose impact can be observed in the long term. As a summative evaluation, the same Likert-scale can be conducted as a delayed survey after the necessary revisions in scoring rubrics or PTs in order to check the impact of the revised version and to gain more insight. The same study can be replicated in experimental research design to find out the effectiveness or any possible superiority of the combination of TA and PBA to TA only or PBA only language learning settings. This research may be assisted by collecting the PT scores of the students in order to check whether there is a sound correlation between their perceptions on the learning outcomes and the corresponding

PT scores. For a more assessment-oriented research, a study can be designed to collect the traditional exam results in order to investigate if PTs have any impact on student achievement or not.



## REFERENCES

- Ak, S. (2013). *The evaluation of performance tasks given at primary education according to students' teachers' and parents' opinions*. Master Dissertation, Çağ University, Mersin.
- Al-Nouh, N. A., & Taqi, H. A. & Abdul-Kareem, M. M. (2014). EFL Primary School Teachers' Attitudes, Knowledge and Skills in Alternative Assessment. *International Education Studies*, 7(5), 68-84.
- Alderson, J.C. (1991). Language testing in the 90's: How far have we come? How much further have we to go? In Anivan, S. (eds.), *Current developments in language testing*. Singapore: SEAMEO Regional Language Center.
- Alderson, J.C., & Banerjee, J. (2001). *Language testing and assessment (Part I)*. Lancaster University, UK, Lang.Teach. Cambridge University Press. doi: 10.1017/S0261444801001707
- Assessment, Articulation, and Accountability. (1999). A Foreign Language Project Funded by a Federal Grant from the Foreign Language Assistance Program. North Carolina Department of Public Instruction, Second Languages. Retrieved from <http://www.ncpublicschools.org/docs/curriculum/worldlanguages/resources/aa/aaa.pdf>
- Bachelor, R. B. (2015). *Alternative Assessments and Student Perceptions in the Foreign Language Classroom* (Ed.D.) Dissertations. Paper 74.
- Bachman, L. F. (1991). What Does Language Testing Have to Offer? University of California, Los Angeles. *TESOL Quarterly*, 25 (4), 671-704.
- Bachman, L. B., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bailey, K. M. (1998). *Learning about language assessment: dilemmas, decisions, and directions*. US, Heinle & Heinle.
- Bamford, J., & Day, R. R. (2003). *Extensive Reading Activities for Teaching Language*. United Kingdom, Cambridge University Press.
- Bayram, E. (2015). *An investigation of language assessment preferences of pre-service EFL teachers and underlying factors*. Master dissertation, Karadeniz Technical University, Trabzon.
- Brooks, L. A. (1999). *Adult ESL student attitudes towards performance-based assessment*. Master dissertation, Department of Curriculum, Teaching and Learning Ontario Institute for Studies in Education of the University of Toronto.

- Brown, G. T. L. (2001–2003). Teachers' conceptions of assessment (TCoA) inventory (versions 1–3). Unpublished test. Auckland, NZ: University of Auckland.
- Brown, H. D. (2004). *Language assessment principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. (2nd ed). White Plains, New York: Pearson Longman.
- Burnaz, Y. E. (2011). *Perceptions of EFL learners towards portfolios as a method of alternative assessment: A case study at a Turkish state university*. Master dissertation, Middle East Technical University, Ankara.
- Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *USA Language Testing*, 27(1), 5–31. Retrieved in 2018 from www.sagepub.co.uk. DOI:10.1177/0265532209346370.
- Büyükduman, F. İ. (n.d.) *Alternative Assessment Tools in ELT*. Ph.D. dissertation, Istanbul Sehir University, İstanbul. Retrieved from <http://www.academia.edu>. doi:10667021/Alternative\_Assessment\_Tools\_in\_ELT
- Büyükkarci, K. (2010). *The effect of formative assessment on learners' test anxiety and assessment preferences in EFL context*. Ph.D. dissertation, Çukurova University, Adana.
- Büyüköztürk, Ş. (2002). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum*. Ankara: Pegem A Yayıncılık.
- Can, E. (2017). *English teachers' classroom assessment practices and the ethics of classroom assessment practices*. Master dissertation, Çağ University, Mersin.
- Canagarajah, S. (2006). Changing Communicative Needs, Revised Assessment Objectives: Testing English as an International Language. *Language Assessment Quarterly*, 3(3), 229–242.
- Carless, D. (2011). From testing to productive student learning: Implementing formative assessment in Confucian-Heritage settings. *Journal of Second Language Teaching and Research*, 1 (2), 155 – 158.
- Celen, K. M. (2016). *Program evaluation of an English language teacher education practicum: Insights from supervisors, student teachers, and graduates*. Master dissertation, Boğaziçi University, İstanbul.
- Ceyhan, E., & Gürcan-Namlu, A. G. (2000). Bilgisayar kaygısı ölçeği: Geçerlik ve güvenilirlik çalışması [The Computer Anxiety Scale: Validity and reliability]. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi*, 10(2), 77–93.
- Chan, Y-C. (2008) Elementary school EFL teachers' beliefs and practices of multiple assessments. *National Taipei University of Education Reflections on English Language Teaching*, 7(1), 37–62.

- Chen, Y. H. (2007). *Exploring the Assessment Aspect of Differentiated Instruction: College EFL Learners' Perspectives on Tiered Performance Tasks*. University of New Orleans Theses and Dissertations. (UMI No.837).
- Chinda, B. (2014). Teachers' Reactions Towards Performance-based Language Assessment. *PASAA*, 48, 57-88.
- Cirit, N. C. (2015). Assessing ELT Pre-Service Teachers via Web 2.0 Tools: Perceptions toward Traditional, Online and Alternative Assessment. *The Turkish Online Journal of Educational Technology*, 14(3), 9-19.
- Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, 20, 147–161.
- Cohen, A. D. (1995). *Assessing language ability in the classroom* (2nd ed.). Boston, MA: Heinle and Heinle.
- Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics* 29, 90–100.
- Çeliker, G. (2015). *Eğitim bilimleri ve öğretmen yetiştirme alan uzmanlarının eğitimde program değerlendirme öz-yeterlik düzeylerinin incelenmesi*. Master dissertation, Eskişehir Osmangazi University, Eskişehir.
- Çetin Bethard, L. M. (2011). *An investigation into the implementation of alternative assessment in the young learner classroom*. Ph. D. Dissertation, Middle East Technical University, Ankara.
- Daloglu, A., & Coşkun, A. (2010). Evaluating an English Language Teacher Education Program through Peacock's Model. *Australian Journal of Teacher Education*, 35(6), 24-42.
- Dittmann-Domenichini, N., Halbherr, T., & Schlienger, C. (2014). *Assessing assessments: student evaluation of online and paper-based examinations at ETH Zurich*. Swiss Federal Institute of Technology Zurich (ETH), Center for Educational Development and Technology.
- Doğan, A. C. (2009). *Instructors' Views Regarding Assessment Tools*. Master dissertation, Hacettepe University, Ankara.
- Doğan, C. D., & Kutlu, Ö. (2010). Factors effecting level of preferring alternative assessment methods. *ScienceDirect, Procedia Social and Behavioral Sciences*, 9, 1135–1139.
- Farhady, H. (2003). Classroom Assessment: A Plea for Change. *Iran University of Science and Technology*, 1-13. Retrieved from [www.researchgate.net](http://www.researchgate.net).
- Farhady, H. (2005). Language Assessment: A Linguametric Perspective. *Language Assessment Quarterly*, 2(2), 147-164. Retrieved from <https://www.researchgate>. DOI:10.1207/s15434311laq0202\_3



- Fitzpatrick, R., & Morrison, E. J. (1971). *Performance and product evaluation*. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 237-270). Washington, DC: American Council on Education.
- Gaies, S. J. (1992). *An Approach to the Evaluation of ELT Preparation Programs*. University of Northern Iowa. Retrieved from <https://files.eric.ed.gov/fulltext/ED369276.pdf>
- Garci'a, G. E., & Pearson, P. D. (1994). Assessment and Diversity. *Review of research in education*, 20(1), 337-391.
- Göçtü, R. (2013). *Comparison of traditional and alternative assessment in English Language Teaching in high schools (Georgian case)*. Ph.D. dissertation, International Black Sea University, Tbilisi.
- Gökçen, R. A. (2005). *Instructors' and administrators' attitudes towards project work as an alternative assessment tool and as an instructional approach at Karadeniz Technical University School of Foreign Languages Department of Basic English*. Master dissertation, Bilkent University, Ankara.
- Gönen, K. (2013). *Teachers' perceptions of classroom based language assessment in tertiary level English language programs in Turkey*. Master dissertation, Çağ University, Mersin.
- Gönül, E. (2010). *6. Sınıf öğrencilerinin, öğretmenlerinin ve velilerinin performans görevleri hakkında görüşleri*. Master dissertation, Ankara University, Ankara.
- Grabin, L. A. (2007). *Alternative assessment in the teaching of English as a foreign language in Israel (Doctoral Dissertation)*. University of South Africa, South Africa.
- Gunn, C. (1999). They Love It, but Do They Learn from It? Evaluating the Educational Impact of Innovations, *Higher Education Research & Development*, 18(2), 185-199, doi: 10.1080/0729436990180203.
- Gürsoy, S. (2013). *The English proficiency exam in EFL context: A validation study*. Master dissertation, Çağ University, Mersin.
- Hacısalıhoğlu, M. (2013). *6. Sınıf öğrencilerinin proje ve performans görevleriyle ilgili görüşleri ve karşılaştıkları güçlükler*. Master dissertation, Karadeniz Technical University, Trabzon.
- Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins and C. Davison (Eds), *The International handbook of English language teaching*, 1, 487-504. Norwell, MA: Springer.
- Han, T. (2014). Turkish EFL Teachers' Assessment Preferences and Practices in the Context of Constructivist Instruction. *Journal of Studies in Education*, 4(1), 77-93.

- Herman, J. L. (1992). *A Practical Guide to Alternative Assessment*. Association for Supervision and Curriculum Development, Alexandria, Va. Office of Educational Research and Improvement (ED), Washington, DC. ISBN-0-87120-197-69.
- Howard-Williams, D., Tomalin, M., Watcyn-Jones, P. & Woods, E. (2001). *Instant lessons (1) (2), and (3) (elementary, immediate, and advance)*. Harlow: Longman.
- Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5(1), 8-11.
- Huerta-Macias, A. (2002). *Assessment Methodology in Language Teaching*. An Anthology of Current Practice Edited by Jack C. Richards Willy A. Renandya Cambridge University Press.
- Kadidja, K. (2015). *The Impact of Performance-Based Assessment on University ESL Learners' Motivation*. Master dissertation, Minnesota State University, Mankato. All Theses, Dissertations, and Other Capstone Projects. (UMI No. 402). Retrieved from DOI: <https://cornerstone.lib.mnsu.edu/etds/402>.
- Kavaliauskienė, G., & Anusienė, L. (2007). The challenges for ESP learners: Alternative assessment of performance and usefulness of class activities. *SOCIALINIS DARBAS*, 6(1), 134-144.
- Kızılkaya, B. (2014). *Students' attitudes towards alternative assessment: A survey-based study on alternative assessment*. Master dissertation, İstanbul University, İstanbul.
- Law, B., & Eckes, M. (1995). *Assessment and ESL*. Peguis publishers: Manitoba, Canada.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*, 20(8), 15-21. doi: 10.3102/0013189X020008015.
- Long, M. H. (1984). Process and Product in ESL Program Evaluation. *TESOL Quarterly*, 18(3), 409-425.
- Mammadov, S. (2015). The Education of gifted K-8 students in Turkey: Policy analysis and program evaluation. *Roeper Review*, 37, 139-149. Retrieved from <https://doi.org/10.1080/02783193.2015.1047548>.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Mid-Continent Regional Educational Lab., Aurora, CO. ISBN-0-87120-225-5.

- McMillan, J. H., & Workman, D. J. (1998). *Classroom Assessment and Grading Practices: A Review of The Literature*. Virginia Commonwealth University. Metropolitan Educational Research Consortium, Richmond, Va. Retrieved from <https://files.eric.ed.gov/fulltext/ED453263.pdf>
- Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments. *Review of Educational Research*, 3 (62), 229-258.
- Metin, M. (2013). Teachers' difficulties in preparation and implementation of performance task. *Educational Sciences: Theory & Practice*, 13(3), 1664-1673. DOI:10.12738/estp.2013.3.1452
- Mızıkacı, F. (2006). A systems approach to program evaluation model for quality in higher education. *Quality Assurance in Education*, 14 (1), 37-53.
- Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools*. New York: The Free Press.
- Moon, T. R., Callahan, C. M., Brighton, C. M., & Tomlinson, C. A. (2002). *Development of Differentiated Performance Assessment Tasks for Middle School Classrooms*.
- Moss, P. A. (1992). Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. *Review of Educational Research*, 3 (62), 229-258.
- Nasab, F. G. (2015). Alternative versus Traditional Assessment. *Journal of Applied Linguistics and Language Research*, 2 (6), 165-178.
- Norris, j. (2009). Understanding and improving language education through program evaluation: Introduction to the special issue. *Language Teaching Research*, 13(1), 7-13.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge, UK: Cambridge University Press.
- Ornstein, A. C., & Hunkins, F.P. (1998). *Curriculum: Foundations, Principles and Issues*. Englawood Cliffs, NJ, Prentice Hall.
- Owen, J. M. (2007). *Program evaluation: Forms and approaches* (3rd ed.). New York: The Guilford Press.
- Önal, A. (2010). *Testing and assessment of speaking skills in preparatory classes*. Master dissertation, Selçuk University, Konya.
- Özeren, E. (2013). *Alternative, a Meta-analysis Study on Measurement and Evaluation Tools*. Master dissertation, Fırat Uuniversity, Elazığ.
- Paksoy, M. (2015). *Ortaokul öğrencilerinin performans görevlerinde bilişim etik kurallarına uyma durumları*. Master dissertation, Pamukkale University, Denizli.

- Peacock, M. (2009). The evaluation of foreign-language-teacher education programmes. *Language Teaching Research*, 13(3), 259-78.
- Perlman, C. (2003). Performance Assessment: Designing Appropriate Performance. U.S. Department of Education.
- Pickering, K. (2012). *Communicate 1 / 2 Listening and Speaking skills book*. MacMillan Publishing.
- Pulat A. E. (2014). *Performans görevlerinin öğrencilerin eleştirel düşünme eğilimlerine ve akademik başarılarına etkisi*. Master dissertation, Ankara University, Ankara.
- Richards J. C., & Renandya, W. A. (2002). *Methodology in Language Teaching An Anthology of Current Practice*. Cambridge: Cambridge University Press.
- Roberts, J. E. (2008). Developing Rubrics. Northern Virginia Community College, the US. Retrieved , 2018 from <http://www.nvcc.edu/assessment/docs/PS6.developingrubrics>
- Sönmez, T. (2013). *The Effects of Formative Assessment on the Autonomy of Turkish Adult EFL Learners*. Master disssertation, Necmettin Erbakan University, Konya.
- Spector-Cohen, E. (2007). Integrating Performance Assessment in the EAP Classroom. *The Internet TESL Journal*, 3 (13). Retrieved, 2018 from <http://iteslj.org/Articles/Spector-Cohen-EAP.html>.
- Stipek, D. (2002). *Motivation to learn: Integrating theory and practice*. Needham Heights, MA, Allyn and Bacon.
- Streff, R. J. (2016). *A Qualitative Case Study of Strategies for Choosing and Evaluating Alternative Assessments in Online Higher Education*. Ph.D. dissertation, Walden University Walden, the U.S.
- Struyven, K. Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Katrien University of Leuven (KULeuven), Belgium Assessment & Evaluation in Higher Education*, 30(4), 331-347.
- Stufflebeam, D. (2001). Evaluation Models. *New Directions for Evaluation*, 7(89),7-97.
- Taş, B. (2016). *Assessment preferences and test anxieties of foreign and Turkish university students in English preparatory classes*. Master dissertation, Süleyman Demirel University, Isparta.
- Taylor, L. (2006). The changing landscape of English: implications for language assessment. *ELT Journal*, 60(1), 51-60. doi:10.1093/elt/cci081

- Thornbury, S. (2006). *An A-Z of ELT: A dictionary of terms and concepts used in English Language Teaching*. Macmillan Publishers Limited.
- Topkaya, E., & Küçük, Ö. (2010). An evaluation of 4th and 5th grade English language teaching program. *Elementary Education Online*, 9(1), 52-65. Retrieved from <http://ilkogretim-online.org.tr>.
- Toumi Lafi. (n.d.). Alternatives in Assessment. Retrieved from [http://www.academia.edu/25264065/ALTERNATIVES\\_IN\\_ASSESSMENT](http://www.academia.edu/25264065/ALTERNATIVES_IN_ASSESSMENT)
- Watcyn-Jones, P. (2001). *Vocabulary Games and Activities 1*. Harlow : Penguin English, Penguin English photocopyables.
- Watcyn-Jones, P., & Howard-Williams, D. (2001). *Grammar Games and Activities 1*. Harlow : Penguin English, Penguin English photocopyables.
- Weigle, S. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weiss, C. H. (1998). *Evaluation: methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 7033-713.
- Wiggins, G. (1990). The case for authentic assessment. ERIC Digest. ERIC Document Reproduction Service No. ED 328 611. Practical Research, Assessment & Evaluation, 2(2). Retrieved from <http://PAREonline.net/getvn.asp?v=2&n=2>.
- Worthen, B. R (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, 74(6), 444-456.
- Yang, W. (2007). *Evaluation of a Induction Program in a U.S. University English Language Institute: Towards Useful Evaluation*. Master dissertation, University of Hawai'i. doi: 10.1.1.995.465&rep=rep1&type=pdf
- Yavuz, A., & Topkaya, E. (2013). Teacher educators' evaluation of the English language teaching program: A Turkish case. *Research on Youth and Language*, 7(1), 64-83.
- Yildirim, 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.
- Yoon-Hee,N., Byung-Kyoo, A., & Hyung-Sun, K. (2008). Evaluating an in-service English teacher training program from multiple perspectives. *English Teaching*, 63(4), 273-302.
- Yüce, Z. (2015). *Pre-service English language teachers' conceptions of assessment and assessment practices*. Master dissertation, Çağ University, Mersin.

Zaimođlu, S. (2013). *Teachers' and students' conceptions of assessment in a university EFL preparatory school context*. Master dissertation, ađ University, Mersin.



## APPENDICES

### APPENDIX A Guiding Documents for students

#### Guide-chart for Extensive Reading PTs

Here is a GUIDE-CHART that shows how you should choose your graded reader according to YOUR LEVEL. It suggests you how the level of your readers should get higher.

	ELEMENTARY	PRE-INTERMEDIATE	INTERMEDIATE
TERM 1 EXTENSIVE READING EVALUATION 1	Starter-A1	A2-B1	B1
TERM 2 EXTENSIVE READING EVALUATION 2			
EXTENSIVE READING EVALUATION 3	A2-B1	B1	B1-B2
EXTENSIVE READING EVALUATION 4			

#### Warning notice for students before PT deadlines

SEVGİLİ ÖĞRENCİLER,

MID-TERM HAFTASI SONRASI 27 KASIM 2017 PAZARTESİ LISTENING/SPEAKING & GRAMMAR & VOCABULARY PERFORMANS GÖREVLERİNİ O DERSİ VEREN OKUTMANLARA BİLDİRMEK İÇİN SON TARİHTİR.

PERFORMANS GÖREVLERİ SÜREÇ TAKİBİ GEREKTİRİR. HER HAFTA ÖĞRENDİĞİNİZ KONU & KELİMELERİ GÖREVİNİZE EKLEYEREK YÜRÜTMENİZ TAVSİYE EDİLİR. BÖYLECE TÜM GÖREVLERİ SON HAFTAYA BIRAKMAMIŞ OLURSUNUZ.

PERFORMANS GÖREVLERİNİN İÇERİĞİ 11. HAFTA & 14. HAFTA DERSLERDE İŞLENEN KONULARI KAPSAMAKTADIR. BU SEBEPLE BU HAFTALARDA DEVAMSIZLIK YAPMAMAYA ÖZEN GÖSTERİNİZ.

**Warning notice for the presentation week**

SEVGİLİ ÖĞRENCİLER,

14. HAFTA CUMA GÜNÜ 22 ARALIK 2017 TASKLARIN SON TESLİM TARİHİDİR.

15. HAFTA (25-29 Aralık 2017) DERS PROGRAMINIZDA KIRMIZI İLE İŞARETLİ OLAN TÜM DERSLERDE SINIFTA BULUNMANIZ ZORUNLUDUR. HAFTA BOYUNCA TÜM DEVAMSIZLIKLAR BELİRLENEN DERSLERDE ALINACAKTIR.

TASKINI TESLİM EDİP BELİRLENEN DERSLERDE SINIFA GELMEYEN ÖĞRENCİNİN NOTU KIRILACAKTIR.

DEĞERLENDİRME HAFTASINDA ÖĞRENCİLER HAZIRLADIKLARI TASKLARI SINIFTA SUNACAKLARDIR. (OKUTMAN & ÖĞRENCİLER TASKI HAZIRLAYAN ÖĞRENCİYE HAZIRLIK AŞAMASI VE TASK İLE İLGİLİ GENEL SORULAR SORABİLİR.)

NOT: DERS PROGRAMINDAKİ BOŞ SAATLERDE SINIFTA GÜRÜLTÜ YAPILMAMASI GEREKMEKTEDİR. BOŞ SAATLER SELF-ACCESSTE (B BLOK ÜST KAT) DEĞERLENDİRİLEBİLİR.



## APPENDIX B      Guiding Documents for instructors

### Information About Vocabulary PT

- Students have to prepare a vocabulary task, one for each term. The aim is to prepare tasks accordingly to their knowledge of vocabulary and its usage. **The tasks must cover the vocabulary studied between weeks 10- 14. Announce the chapter numbers to the students.**
- The tasks have been categorized according to students' interests and abilities. The tasks are the same for every level.
- **Students will be informed about the tasks on week 8** and they are expected **to decide on their task by week 11 on Monday at the latest.** So any decision made between weeks 9-11 will be accepted.
- Once they have decided, they cannot change the task. They will start doing their tasks between weeks 10-14.
- Make sure you inform your students that they must choose a different task for each term. Teachers must note down the tasks for each term on the form given. Keep the form for both school terms. (Task Distribution Form)
- The scores (marks) the students get from vocabulary tasks (as in all skills) constitute % 4 of their GPA.
- **Students will hand in their tasks on week 14 on Friday at the latest.**  
**NOTE:** Week 14 is also "FILM WEEK" for vocabulary lessons. If you have vocabulary lessons on Friday, you can accept the tasks during break time or before. No delays should be tolerated and marked.
- The process of the tasks should be recorded on a CD. Only CD's are accepted.
- **The tasks will be assessed on week 15 during office time.**
- Also 2 lesson hours should be allocated to assess (oral checking) the students in class. Students must be present in those 2 hours or they will be considered absent for the whole week.
- During these 2 hour lessons, multiple questions can be asked to the students about the preparation phase of their tasks. This is recommended to prevent forgery. If you are not sure of the task being done by the student, no marking is given. Weekly Schedule to follow is given for Vocabulary PTs below:

<b>WEEK 8</b>	STUDENTS ARE GIVEN INFORMATION ABOUT THE (PT) TASKS
<b>WEEK 9-11</b>	DECISION OF THE TASK NUMBER
<b>WEEK 10-14</b>	DOING THEIR TASKS
<b>WEEK 14</b>	DUE DATE (on Friday the latest)
<b>WEEK 15</b>	ASSESSMENT WEEK

**Process and Product-oriented guiding questions (for instructors)**

1. How was the preparation process of your PT?
2. What kinds of challenges have you experienced? (If any)
3. Did you benefit from any resources (books, magazines, the internet...etc.) while preparing your PT?
4. What kind of materials did you use to prepare your PT?
4. Did you enjoy while preparing your PT?
5. How did you find your product? After completing your task, did you notice that there is something you should change in your PT?
6. How many hours or days did it take to complete your PT?
7. Did you add any creativity or extra visuals to your PT different from the written instructions?
8. Did you show your product to any of classmates and get feedback from her/him?
9. After completing your PT, did you do any self-checking?
10. Do you suggest your PT to your classmates? Why/Why not?

**Feedback Samples (for instructors)**

Your PT is a nice piece of work, but there is a missing part  
.....

I liked your PT a lot. It is an excellent job! Well done!

I loved the way you used your creativity in your PT, but it gives only basic information about.....

Maybe it is a better idea to  
.....

It seems quite original, but I think you have misunderstood  
.....

I guess it is one of the best examples of this task!

Next time you had better ask for more guidance because  
.....

\*If you are not satisfied with the PT presentation of the student, you may ask some specific questions about his/her PT.

## APPENDIX C

### Sample Schedule in PT Assessment and Evaluation Week (Week 15)

1	<u>LSP</u> <u>Aygül Ş</u>	R Sezen B	<u>R</u> <u>Sezen B</u>	LSP Aygül Ş	<u>Y</u> <u>Sevim E</u>
2	<u>LSP</u> <u>Aygül Ş</u>	<u>WR</u> <u>Güner M</u>	<u>R</u> <u>Sezen B</u>	LSP Aygül Ş	<u>Y</u> <u>Sevim E</u>
3	GR Fatma K	<u>WR</u> <u>Güner M</u>	GR Fatma K	WR Güner M	WR Güner M
4	GR Fatma K	R Sezen B	GR Fatma K	WR Güner M	GR Fatma K
5		LSP Aygül Ş	LSP Aygül Ş		
6	R Sezen B			<u>GR</u> <u>Fatma K</u>	
7	LSP Aygül Ş			<u>GR</u> <u>Fatma K</u>	

## APPENDIX D      Samples of Performance-based Assessment Tasks

### **Extensive Reading PTs**

#### **PT1 Activity Description: Writing a report**

Learning objectives: Reading for pleasure, information and general understanding, critical thinking

Number of participants: Individual work

Instructions: Choose a text that is suitable for your language proficiency level. It can be a short story or an article from newspapers or magazines. Read it and then prepare a report on the basis of the following issues:

- 1) Why did you choose this text? Justify your answer or list the reasons why you choose this text.
- 2) What does the text mainly discuss? (Between 50 and 100 words)
- 3) What is the most interesting information or message you've learned from the text?
- 4) Comment on the title. Do you think it is appropriate for the text? Why? / Why not? If not, what title would you suggest?
- 5) Write your comments and reactions to the text. Would you recommend anybody to read the text you have read? If yes, to whom and why? If no, state your reasons.

#### **PT3 Activity Description: Preparing a quiz show**

Learning objectives: Describing, strategic thinking, classifying, sequencing, identifying cause and effect

Number of participants: Individual work

Instructions: Choose a graded reader that is suitable for your language proficiency level.

Read it and then prepare a quiz show about the story. You are the quiz show presenter. You are going to write various types of questions in the show and provide the right answers as well. You will prepare it like a poster.

The types of questions may include these:

- True-False questions
- Who said what. Match the sentences from the story to the characters.
- Write an ordered timeline of events in the story in one line sentences. Jumble them up and the aim is to re-order them.
- Giving distinguishing adjectives that describe characters from the book (stubborn, daring...) and guessing who it is.

(Adapted from Bamford and Day, 2003)

## **Vocabulary PTs**

### **PT: 10**

**Activity description:** Who wants to be a millionaire?

**Learning Objectives:** To activate target vocabulary by means of preparing a set of general knowledge questions on a TV show

**Number of participants:** This is an individual task.

**Instructions:** The student prepares a quiz show in which he/she asks 15 questions. The target words could be involved in either the questions or in the answers. The student will be the presenter of the quiz show and s/he could get help from one of his/her friends to be a contestant. S/He will record the quiz show process on a video on CD.

### **PT: 11**

**Activity description:** An interview with a future colleague

**Learning Objectives:** To reinforce target vocabulary through an interview

**Number of participants:** This is an individual task.

**Instructions:** The student finds a person who has already started to professionalize in the same field as him/her. S/he prepares at least 15 questions including the target vocabulary to learn about his or her future job. The Q/A (question-answer) process will be recorded on video on CD.

### **PT: 13**

**Activity description:** To prepare a presentation comparing two cultures

**Learning Objectives:** To make use of the target vocabulary by describing two distinct cultures

**Number of participants:** This is an individual task.

**Instructions:** The student presents his speech via PPT. He/she should use visuals and compare the cultures of two different countries using 15 of the target words. (Turkish culture is not necessary.) In the presentation, there should be certain categories included such as people, life styles, food, nature, economy, education and so on. The student can use video recordings in the presentation as extra visual materials. The presentation will be recorded on CD.

(Developed by the vocabulary committee adapting from Vocabulary Games and Activities (Watcyn-Jones, 2001), and Instant Lessons (Howard-Williams; Tomalin; Watcyn-Jones & Woods, 2001)

## Grammar PTs

### ELEMENTARY LEVEL GRAMMAR PTs

#### PT 1

**Activity Description:** Crime scene report

**Learning Objectives:** To practice ► **can-could-be able to-may-might-shall**, (5 sentences)

► **relative clauses/ relative pronouns**, (5 sentences)

► **reflexive-emphatic pronouns/which?/one-ones**

(5 sentences)

To present ideas according to some specific data

**Number of Participants:** Individual work

**Instructions:** You are a detective who is trying to solve a case which may be a murder or a suicide. Take 5 photos of the evidence you have collected from the crime scene. By using the clues present your view about the case. Shoot a video of yourself and the photos.

!!!The clues you need are on “some!” of the walls in the school.

#### PT 3

**Activity Description:** Dubbing

**Learning Objectives:** To practice ► **can-could-be able to-may-might-shall**, (5 sentences)

► **infinitive/-ing form/too-enough**, (5 sentences)

► **reflexive-emphatic pronouns/which?/one-ones**

(5 sentences)

To use specific structures in a defined context, to use appropriate intonation and stress

**Number of Participants:** Individual work

**Instructions:** The following links will direct you to short movies *Piper* and *Invisible*. Choose one of them and record your voice as if you are dubbing.

► <https://www.youtube.com/watch?v=QVGeilNsJFU>

► <https://www.youtube.com/watch?v=-N5Dy4R1Fco>

(Adapted from [www.teachingenglish.org.uk](http://www.teachingenglish.org.uk) and Grammar Games and Activities Watcyn-Jones & Howard-Williams, 2001)

## **Listening / Speaking PTs**

### **PRE-INTERMEDIATE LEVEL**

#### **PAIR WORK**

1. Find a pair until the end of week 9 and inform your teacher.
2. You are going to shoot one short movie about either the topics of week 11 or week 12. You and your pair will decide the unit that you want to work on and inform your teacher until the end of week 11.
3. The video will last 5 to 10 minutes and has to include the target vocabulary and expressions that the course book covers.
4. The video will be handed in until the end of week 14. The ones who pass the deadlines will not be evaluated and will fail by receiving zero.
5. Videos can be sent via e-mail or given as CD. The transcripts of the both videos and the extra materials such as the menu or the logo of the company will be added to the mail or CD.

**TASK 1: EATING OUT:** You are opening a new restaurant.

#### **The first part of the video:**

1. Decide what kind of restaurant it will be. (fast food restaurant/ a fancy restaurant/food truck, etc.)
2. Discuss the food that you want to cook at the restaurant with your pair. Talk about the food that you are going to include in the menu.

\*\*\* Prepare the menu on colorful papers. (You are going to hand in the menu, as well.)

#### **The second part of the video:**

One of you will be the customer and the other will be the waiter. The waiter will welcome the customer and show him/her the menu. The customer orders what he or she likes. There must be a problem either on the food or the bill. Waiter will solve the problem.

**TASK 2: YOUNG WORKERS:** You are opening a new business.

**The first part of the video:**

1. Decide the business sector you want to work in. (tourism, entertainment, transportation, finance, etc.)
2. Decide the name of the company, the logo and the slogan. Discuss your ideas and find the best one.
3. Decide the qualifications of your employees that you need for your business. Discuss and prepare a list of the interview questions according to the qualifications you are looking for.

**The second part of the video:**

One of you will be the job applicant and one of you will be the employer. The employer will make the interview. The job applicant will try to impress the employer. (or you can choose to add humorous details to the job applicant.)

(Adapted from “Communicate 1 / 2” Listening and Speaking Skill course material Pickering, 2012)



## APPENDIX E      Scoring Rubrics for PTs

### Scoring Rubric of Extensive Reading PT

<b>Excellent</b> <b>4 pts.</b>	Task <u>completed excellently</u> . Content is <u>knowledgeable</u> and <u>relevant</u> to assigned topic with <u>artistic creativity</u> / <u>originality</u> .
<b>Good</b> <b>3 pts.</b>	Task <u>completed fully</u> . Content is <u>adequate mostly relevant</u> to topic with <u>some</u> artistic creativity / originality.
<b>Average</b> <b>2 pts.</b>	Task <u>completed partially</u> . Content is <u>limited</u> with <u>insufficient</u> artistic creativity / originality.
<b>Poor</b> <b>1 pts.</b>	Task is <u>not completed</u> . Content is <u>not enough</u> to evaluate with <u>no</u> artistic creativity / originality.

(Developed by following the guidelines in Weigle 2002)

## Scoring Rubric of Grammar PTs

	<b>Hits the Target</b>	<b>Approaches the Target</b>	<b>Below the Target</b>	<b>Misses the Target</b>
<b>Accurate use of Required Structure 1</b> ..... ..... <b>20 points</b>	It includes as many error-free and fluently uttered sentences as stated in the instruction.  <b>(20-15 pts.)</b>	It includes expected number of sentences with reasonable accuracy OR fewer error-free sentences than stated in the instruction. <b>(14-10 pts.)</b>	It includes almost no sentences with the target structure AND / OR many mistakes which lead to comprehension problems. <b>(9-5 pts.)</b>	It doesn't include any sentences with the target structure OR sentences are all flawed.  <b>(4-0 pts.)</b>
<b>Accurate use of Required Structure 2</b> ..... ..... <b>20 points</b>	It includes as any error-free and fluently uttered sentences as stated in the instruction.  <b>(20-15 pts.)</b>	It includes expected number of sentences with reasonable accuracy OR fewer error-free sentences than stated in the instruction. <b>(14-10 pts.)</b>	It includes almost no sentences with the target structure AND / OR many mistakes which lead to comprehension problems. <b>(9-5 pts.)</b>	It doesn't include any sentences with the target structure OR sentences are all flawed.  <b>(4-0 pts.)</b>
<b>Accurate use of Required Structure 3</b> ..... ..... <b>20 points</b>	It includes as many error-free and fluently uttered sentences as stated in the instruction.  <b>(15-11 pts.)</b>	It includes expected number of sentences with reasonable accuracy OR fewer error-free sentences than stated in the instruction. <b>(10-7 pts.)</b>	It includes almost no sentences with the target structure AND / OR many mistakes which lead to comprehension problems. <b>(6-3 pts)</b>	It doesn't include any sentences with the target structure OR sentences are all flawed.  <b>(3-0 pts.)</b>
<b>Task Achievement</b> <b>15 points</b>	It covers all the instructions given in the task.  <b>(15-12 pts.)</b>	It covers most of the instructions given in the task.  <b>(11-8 pts.)</b>	It covers few of the instructions given in the task.  <b>(7-4 pts.)</b>	It covers almost none of the instructions given in the task.  <b>(3-0 pts.)</b>
<b>Content</b> <b>15 points</b>	It is very informative. <b>(15-12 pts.)</b>	It has some information about the topic. <b>(11-8 pts.)</b>	It has limited information about the topic. <b>(7-4 pts.)</b>	It has no information about the topic. <b>( 3-0 pts.)</b>
<b>Organization</b> <b>10 points</b>	It is well-organized and it has smooth progression with a perfect transition from one idea to the next. <b>(10-9 pts.)</b>	It is nearly well-organized with an adequate transition which causes clear understanding.  <b>(8-6 pts.)</b>	It has limited organization with a poor transition between ideas.  <b>(5-3 pts.)</b>	It is poorly organized and demonstrates serious problems with progression of ideas.  <b>(2-0 pts.)</b>

(Developed by following the guidelines in Weigle 2002)

## Scoring Rubric of Listening / Speaking PT

<b>Sentence Structure</b>	_____ 10pts. Very effective and varied use of structures with <b>no or very few noticeable grammar errors</b>	_____ 7pts. Effective and varied use of structures, <b>few grammatical errors but do not obscure the meaning</b>	_____ 4 pts. Ineffective, mostly repetitive, and less varied structures than expected at this level, <b>frequent use of grammatical errors that occasionally obscure the meaning</b>	_____ 1 pt. Use of very basic structures <b>full of grammatical errors that obscure the meaning</b>
<b>Vocabulary</b>	_____ 10 pts. Accurate, appropriate, and varied use of old and newly learnt vocabulary with <b>no or very few noticeable errors</b>	_____ 7 pts. Both old and newly learnt vocabulary is used with <b>some inappropriate words, but do not affect communication</b>	_____ 4 pts. Vocabulary is mostly repetitive <b>and mostly inappropriate, which sometimes affect communication</b>	_____ 1 pt. Very limited vocabulary <b>full of inappropriate words that make understanding impossible</b>
<b>Fluency</b>	_____ 10 pts. Very fluent and smooth speech with <b>no or minimal hesitation</b>	_____ 7 pts. Fluent speech with <b>some noticeable hesitations, but do not disturb the listener</b>	_____ 4 pts. Hesitant speech <b>often results in considerable disturbance to the listener</b>	_____ 1 pt. No connected speech, <b>hesitations prevent the understanding completely</b>
<b>Pronunciation /Intelligibility</b>	_____ 10 pts. Clear speech with <b>no or very few mispronounced words and no or very few stress or intonation errors</b>	_____ 7 pts. Easily understandable speech; <b>some mispronounced words and some stress or intonation errors attract the listener's attention but do not affect understanding</b>	_____ 4 pts. Speech is difficult to understand <b>due to frequent mispronounced words and frequent stress or intonation errors which demand the listener's attention and effort</b>	_____ 1 pt. Totally unclear speech <b>due to incorrect pronunciation, stress, and intonation make understanding very difficult or impossible</b>
<b>Task achievement</b>	_____ 10 pts. Relevant content to the topic, task completed fully with <b>all necessary newly learnt communication strategies</b>	_____ 7 pts. Mostly relevant content to the topic, and task completed almost perfectly with <b>some of necessary newly learnt communication strategies</b>	_____ 4 pts. A little relevant content to the topic, and task completed almost adequately with <b>few and repetitive newly learnt communication strategies</b>	_____ 1 pt. Irrelevant content to the topic, or task completed <b>inadequately due to limited communication strategies and relies heavily on conversation partner to sustain conversation</b>

(Adapted from “Common European Framework speaking evaluation rubric”)

**Scoring Rubric of Vocabulary PT**

<b>Pts.</b>	<b>ACCURATE USE</b>	<b>PRESENTATION</b>	<b>CREATIVITY</b>	<b>TASK COMPLETION</b>
<b>20-25</b>	Accurate and appropriate use of all target words.	The task is very careful and neatly done. It is easy to understand. (An eye-popper!)	A lot of thought was put into making the task different and interesting.	Task completed fully. All of the requirements have been met.
<b>12-19</b>	Students can use the target vocabulary accurately. (with few noticeable errors)	The task is fairly neat and understandable.	Some thought was put into making the task different and interesting.	Task completed almost adequately. Most of the requirements have been met.
<b>5-11</b>	Students have frequent mistakes in the use of target vocabulary.	The task is slightly lacking neatness. Some parts are difficult to understand.	Thought was put into making the task but has not made it different or interesting.	Task completed inadequately. Some of the requirements have been met.
<b>0-4</b>	The task has no relevance and does not use the target vocabulary.	The activity is lacking neatness.	Little thought was put into making the task.	Task not completed. None of the requirements have been met.

Note: If the student has prepared an irrelevant task, assign 10 points.

(Developed by the vocabulary committee following the guidelines by Roberts, J. E. 2008)

## APPENDIX F

### Sample Pages of Assessment PPP of preparatory school

#### PERFORMANSA YÖNELİK ÖLÇME & DEĞERLENDİRME UYGULAMALARI

- ▶ **GENEL AMAÇ:** GELENEKSEL SINAVLARA EK OLARAK ÖLÇME DEĞERLENDİRME ÇEŞİTLİLİĞİ SAĞLAYARAK ÖĞRENCİNİN YABANCI DİL YETKİNLİĞİ, DİLİ KULLANMA BECERİSİ VE DİL ÖĞRENME STRATEJİLERİNİ ÖLÇMEKTİR.
- ▶ **HEDEFLER:** ÖZELLİKLE SINIF DIŞI UYGULAMALAR İLE ÖĞRENCİNİN YABANCI DİLİ SINIF ORTAMI DIŞINDA VE GÜNLÜK AKTİVİTELERDE KULLANMASI HEDEFLENMİŞTİR.
- ▶ **ÖĞRENCİLERİN SINIF İÇİNDE ÖĞRENDİKLERİ DİLBİLGİSİ KALIPLARINI, KELİME KULLANIMLARINI VE EDİNDİKLERİ DİL BECERİLERİNİ AKTİF HALE GETİREBİLMELERİ İÇİN YABANCI DİLDE ÜRETİM YAPMALARI HEDEFLENMİŞTİR.**
- ▶ **TASK ÇEŞİTLERİ:** VİDEO ÇEKME, SUNUM YAPMA, POSTER HAZIRLAMA, OYUN ÜRETME...VB.

#### PERFORMANSA YÖNELİK ÖLÇME & DEĞERLENDİRME UYGULAMALARI

- ▶ **TASKLAR:** ÖĞRENCİNİN SEVİYESİNE VE İLGİ ALANLARINA UYGUNDUR. SEÇENEKLER ARASINDAN ÖĞRENCİ HER DERS İÇİN DÖNEMDE BİR TASK BELİRLER. SADECE OKUMA DERSİ İÇİN İKİ FARKLI TASKA HAZIRLANIR.
- ▶ **TAKİP:** ÖĞRENCİLER TASK TAKİP FORMLARINA SEÇTİKLERİ TASKLARI YAZAR VE İLGİLİ DERSİN HOÇASI BU FORMU TAKİP EDER. TASK SEÇİMİ İÇİN 1 HAFTA SÜRE TANINIR SONRASINDA KESİNLİKLE TASK DEĞİŞİKLİĞİ YAPILAMAZ.
- ▶ **UYGULAMA TARİHLERİ:** OKUMA DERSİ: 5. HAFTA TASK I/II & Kitap belirleme / 9. HAFTA TASK I TESLİM -14. HAFTA TASK II TESLİM  
DİLBİLGİSİ / SÖZCÜK BİLGİSİ / KONUŞMA & DİNLEME BECERİSİ İÇİN 10. HAFTA-14. HAFTA ARASI  
YAZMA DERSİ SINIF İÇİ UYGULAMA OLUP DÖNEME YAYILMIŞTIR.  
**\*DERS PROGRAMINDA 15. HAFTA — BU UYGULAMALARIN DEĞERLENDİRMESİNE AYRILMIŞTIR!**
- ▶ **ORTALAMAYA KATKISI:** HER BİR UYGULAMANIN GENEL ORTALAMAYA DİREKT %4 ORANINDA ETKİSİ VARDIR. BU UYGULAMALAR %20 PERFORMANS BAŞLIĞI ALTINDA TOPLANMIŞTIR.

## **APPENDIX G      Feedback Sample Extract of an instructor**

### **INTERMEDIATE 2 VOCABULARY PT FEEDBACKS**

#### **TASK 4**

**1)Şaban Y.:** Task 4 Memory card game – colourful cards preparation video shooting ok – playing the game part is missing! Task completion mistake

**2) Ece B.:** Task 4 Memory game- I adore the colour of your cards! Fully completed task did you play it with Cem? At last Cem wins the game! The only problem is from presentation you should write on the cards whether it is a definition synonym or antonym to be more understandable (presentation part in rubric)

**3) Rümeyşa:** Task 4 Memory card game fully completed task did you play it with Beyza? At last Rümeyşa wins the game! The only problem is from presentation you should write on the cards whether it is a definition synonym or antonym to be more understandable (presentation part in rubric)

**4) Beyza D.:** Task 4 Memory game- fully completed task did you play it with Rümeyşa? At last Beyza wins the game! The only problem is from presentation you should write on the cards whether it is a definition synonym or antonym to be more understandable (presentation part in rubric)

**5) Nuran:** Task 4 It was a good idea to prepare two videos one for the preparation one for the game. I think your partner İrem does not know English because you directed her all the time. The same mistake for this task presentation part is it a noun verb? You should write the part of speech is it the definition or synonym?

**6) Beyza M.:** Task 4 You played it with your brother did you enjoy the game? You won in the end! I can see only the game not the preparation process and the cards so I broke some points from the presentation & task completion parts.

#### **TASK 2**

**7) Selman Emre-** Task 2 preparing a cross word puzzle on the computer is a good idea practical with little effort in the instructions you are not supposed to play it with a friend but you could try it I only broke some points from the creativity part!

**8) Mertcan T.-** Task 2 You could add a music and I can't see your face no presentation I could recognize you from your bracelet. It was prepared on computer. From the presentation and creativity part I broke some points

**9) Cem T.-**Task 2- Nice oral presentation you have told all the phases of your process. Colorful poster it requires a lot of time & effort. The only problem is about the accurate use of words (in rubric) because you have given the direct dictionary definitions of these words some of them are different from the meanings we have learnt in the class! You can also play it with a friend for more creativity!

**10) Sebile:** Task 2- I expect you to choose a more demanding and creative task. You used the words accurately the content was ok but you did not complete the cross word I cannot see the answers and no video shooting of the process so you got not good points from the presentation and task completion parts in rubric.

## APPENDIX H Student Pictures performing PTs





## APPENDIX I

## Open-ended questionnaires

### Open-ended questionnaire for students

#### CONSENT FORM

Sevgili öğrenci,

Bu araştırmanın amacı performansa dayalı ölçme-değerlendirme yöntemini İngilizce hazırlık okulunda yürütülen öğretim programının bir parçası olarak değerlendirmektir. Araştırma Yeditepe Üniversitesi İngiliz Dili Eğitimi programında yüksek lisans yapmakta olan Nihan Özusu tarafından yürütülmektedir. Bu anketi yanıtlayarak yapacağınız katkı sadece akademik çalışmam için değil aynı zamanda kurumumuz için de önem arz etmektedir. Fikir ve önerilerinizden öğretim programımızı iyileştirmek amacıyla faydalanılacaktır.

Hazırlık sınıfında bu yıl iki yazılı kısa sınav (quiz) uygulamasının biri yerine geçen performansa dayalı uygulamalar düzenlenmektedir. Açık-uçlu iki sorudan oluşan bu anketin amacı uygulama öncesi katılımcılardan ön bilgi toplamaktır.

Bu anketle elde edilen veriler yüksek lisans tez çalışmam için, sadece akademik amaçla kullanılacaktır ve notunuzu etkilemeyecektir. Çalışmadan çıkma hakkınız mevcuttur ve çıkmak istediğinizde bilgileriniz tarafımdan silinecektir. Eğer katılımcı olursanız, yaklaşık 10 dakika ayırmanız gerekmektedir. Çalışma sonuçları ile ilgili bilgi almak isterseniz, benimle aşağıda belirttiğim e-mail adresimden iletişime geçebilirsiniz.

Katkılarınız için şimdiden teşekkürlerimi sunarım.

E-mail: [nihanozuslu@yahoo.com](mailto:nihanozuslu@yahoo.com)

Yukarıdaki bilgilendirmeyi okudum. Bu anket çalışmasında katılımcı olmayı kabul ediyorum.

İMZA:

TARİH:

Sevgili öğrenci,

Lütfen performansa dayalı ölçme-değerlendirme uygulamaları ile ilgili görüşlerinizi belirtiniz.

1. Performansa dayalı ölçme-değerlendirme uygulamalarının pozitif yanları nelerdir?

2. Performansa dayalı ölçme-değerlendirme uygulamalarının negatif yanları nelerdir?

(Developed by the researcher of the present study)



## Open-ended questionnaire for administrators

### CONSENT FORM

Sayın idareci,

Bu araştırmanın amacı performansa dayalı ölçme-değerlendirme yöntemini İngilizce hazırlık okulunda yürütülen öğretim programının bir parçası olarak değerlendirmektir. Araştırma Yeditepe Üniversitesi İngiliz Dili Eğitimi programında yüksek lisans yapmakta olan Nihan Özuslu tarafından yürütülmektedir. Bu anketi yanıtlayarak yapacağınız katkı sadece akademik çalışmam için değil aynı zamanda kurumumuz için de önem arz etmektedir. Fikir ve önerilerinizden öğretim programımızı iyileştirmek amacıyla faydalanılacaktır.

Hazırlık sınıfında bu yıl iki yazılı kısa sınav (quiz) uygulamasının biri yerine geçen performansa dayalı uygulamalar düzenlenmektedir. Açık-uçlu iki sorudan oluşan bu anketin amacı uygulama öncesi katılımcılardan ön bilgi toplamaktır.

Bu anketle elde edilen veriler yüksek lisans tez çalışmam için, sadece akademik amaçla kullanılacaktır. Çalışmadan çıkma hakkınız mevcuttur ve çıkmak istediğinizde bilgileriniz tarafımdan silinecektir. Eğer katılımcı olursanız, yaklaşık 10 dakika ayırmanız gerekmektedir. Çalışma sonuçları ile ilgili bilgi almak isterseniz, benimle şahsen veya aşağıda belirttiğim e-mail adresimden iletişime geçebilirsiniz.

Katkılarınız için şimdiden teşekkürlerimi sunarım.

E-mail: [nihanozuslu@yahoo.com](mailto:nihanozuslu@yahoo.com)

Yukarıdaki bilgilendirmeyi okudum. Bu anket çalışmasında katılımcı olmayı kabul ediyorum.

İMZA:

TARİH:

Lütfen performansa dayalı ölçme-değerlendirme uygulamaları ile ilgili görüşlerinizi belirtiniz.

1. Performansa dayalı ölçme-değerlendirme uygulamalarının size göre olumlu yanları nelerdir?
2. Performansa dayalı ölçme-değerlendirme uygulamalarının size göre olumsuz yanları nelerdir?

(Developed by the researcher of the present study)

## **Open-ended questionnaire for instructors**

### **CONSENT FORM**

Dear colleague,

The purpose of this research is to evaluate performance-based assessment as a component of the language program in this EFL preparatory school. This study is being conducted by Nihan ÖZUSLU, master student in the Graduate School of Education at Yeditepe University. I would be pleased if you could help me by responding a questionnaire. As a researcher, I guarantee that all information you provide in the questionnaires will be treated confidentially. The data collected from you is not only vital for my MA study but also for the improvement of the language program in our institution.

As you know, this year performance-based assessment tasks (PTs) are going to substitute one quiz for grammar, vocabulary, reading, and listening-speaking courses. This open-ended questionnaire, consisting 2 questions, aims to gather information on your preliminary thoughts about PT practices before the implementation. Your answers will be used in my MA thesis only by me for academic purposes. If you have any questions about the questionnaire or the research itself, please feel free to ask me.

The participation in this study is voluntary. If you agree to participate, you will spend approximately 10 minutes responding the items. You have the right to withdraw from the study any time you wish. If you wish to be informed about the study results, you may contact me in person or via my e-mail address below.

Thank you in advance for your contribution.

Email: [nihanozuslu@yahoo.com](mailto:nihanozuslu@yahoo.com)

I have read the information above. I agree to be the participant of this research.

Signature:

Date:

Please comment on the issues below based on your preliminary thoughts.

1. What are the positive aspects of performance-based assessment tasks from your point of view?
2. What are the negative aspects of performance-based assessment tasks from your point of view?

(Developed by the researcher of the present study)

**Likert-scale for administrators****CONSENT FORM**

Sayın idareci,

Bu araştırmanın amacı performansa dayalı ölçme-değerlendirme yöntemini İngilizce hazırlık okulunda yürütülen öğretim programının bir parçası olarak değerlendirmektir. Araştırma Yeditepe Üniversitesi İngiliz Dili Eğitimi programında yüksek lisans yapmakta olan Nihan Özuslu tarafından yürütülmektedir. Bu anketi yanıtlayarak yapacağınız katkı sadece akademik çalışmam için değil aynı zamanda kurumumuz için de önem arz etmektedir. Fikir ve önerilerinizden öğretim programımızı iyileştirmek amacıyla faydalanılacaktır.

2017-2018 Akademik yılında dilbilgisi, kelime bilgisi, okuma ve dinleme-konuşma derslerinde uygulanan performans görevleri klasik kısa sınav (quiz) yerine geçmektedir. Üç ana bölümden oluşan bu anketin amacı performans görevlerini planlama, uygulama, notlandırma, öğrenim kazanımları ve mevcut öğretim programıyla uyumu açısından sizlerden bilgi toplayarak değerlendirmektir.

Bu anketle elde edilen bilgiler yüksek lisans tez çalışmam için, sadece akademik amaçla kullanılacaktır. Çalışmadan çıkma hakkınız mevcuttur ve çıkmak istediğinizde bilgileriniz tarafımdan silinecektir. Eğer katılımcı olursanız, yaklaşık 15 dakika ayırmanız gerekmektedir. Çalışma sonuçları ile ilgili bilgi almak isterseniz, benimle şahsen veya aşağıda belirttiğim e-mail adresimden iletişime geçebilirsiniz.

Katkılarınız için şimdiden teşekkürler.

E-mail: [nihanozuslu@yahoo.com](mailto:nihanozuslu@yahoo.com)

Yukarıdaki bilgilendirmeyi okudum. Bu anket çalışmasında katılımcı olmayı kabul ediyorum.

İMZA:

TARİH:

## DEMOGRAFİK BİLGİ

**BÖLÜM I.** Bu bölüm kişisel bilgilerinizi içermektedir. Lütfen sizin için uygun seçeneğe işaret (✓) koyunuz.

1. Yaşınız: 20-25  26-30  31-35  36-40  41-50  51 ve üzeri

2. Cinsiyet: Erkek  Kadın

3. Eğitim durumu: Lisans  Yüksek lisans  Doktora

4. Mesleki deneyiminiz: 10 yıldan az  10 yıldan çok

5. Sınav hazırlama deneyiminiz oldu mu? EVET  Lütfen açıklayınız \_\_\_\_\_

HAYIR

6. Eğitim hayatınız boyunca hiç ölçme değerlendirme dersi aldınız mı? EVET

HAYIR

7. “Performansa dayalı ölçme değerlendirme” yöntemini bu çalışmadan önce hiç duydunuz mu?

EVET

HAYIR

8. Eğer cevabınız evetse “performansa dayalı ölçme değerlendirme metodunu” nereden biliyordunuz?

Gönüllü okumalar (makale, dergi... vb.)  Yüksek lisans dersleri  Yüksek lisans tezi

Doktora Dersleri  Doktora Tezi  Seminer ve konferanslar  Meslektaşlarım

Çevrimiçi öğretmen forumları  Diğer  Lütfen belirtiniz \_\_\_\_\_

<b>PLANLAMA</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Kararsızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
1. Okutmanlara performans görev uygulamaları öncesi gerekli bilgilendirme sağlandı.	5	4	3	2	1
2. Performans görev uygulamaları öncesi öğrencilere gerekli bilgilendirme sağlandı.	5	4	3	2	1
3. Tüm katılımcılara uygulama öncesi performans sınav yönergeleri (Rubrik) açıklandı.	5	4	3	2	1
4. Öğrenciler performans görevlerini nasıl yapacağını anlamadı.	5	4	3	2	1
5. Öğrencilere Performans görev uygulama komutları ve açıklamaları okutmanlar tarafından net bir şekilde yapıldı.	5	4	3	2	1
6. Performans görevleri için haftalık program önceden duyuruldu.	5	4	3	2	1
7. Performans görev uygulamaları teslim tarihleri öğrencilere duyuruldu.	5	4	3	2	1
8. Öğrenciler PT uygulamaları öncesi sürece hazır hale getirildi.	5	4	3	2	1
<b>UYGULAMA</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Kararsızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
9. Performans görevleri öğrencilerin dil seviyesine uygundu.	5	4	3	2	1
10. Performans görevleri öğrencilerin bilişsel seviyesine uygundu.	5	4	3	2	1
11. Performans görevleri öğrenciler için eğlenceliydi.	5	4	3	2	1

12. Performans görevleri için öğrencilere verilen süre yeterliydi.	5	4	3	2	1
13. Performans görevleri tüm katılımcılar için zaman kaybıydı.	5	4	3	2	1
14. Performans görev sürecinde öğrencilere yeterli rehberlik sağlandı.	5	4	3	2	1
15. Öğrenciler performans görevlerini yaparken sıkıldı.	5	4	3	2	1
16. Performans görevlerini sınıfta sunmak öğrenciler için faydalıydı.	5	4	3	2	1
17. Öğrenciler performans görevlerini zamanında teslim edebildi.	5	4	3	2	1
18. Öğrencilerin çoğu performans görevlerini kendi başlarına hazırladı.	5	4	3	2	1
<b>NOTLANDIRMA</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Kararsızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
19. Performans görevleri belli bir kritere göre değerlendirildi.	5	4	3	2	1
20. Değerlendirme kriterlerinin anlaşılması tüm katılımcılar için kolaydı.	5	4	3	2	1
21. Performans görevleri objektif bir şekilde notlandırıldı.	5	4	3	2	1
22. Performans görev notları öğrenciler için tatmin ediciydi.	5	4	3	2	1
23. Ölçme değerlendirmeye ayrılan 1 haftalık süre yeterliydi.	5	4	3	2	1
24. Performans görevlerinin notlandırılması adil değildi.	5	4	3	2	1

25. Performans görevleriyle ilgili öğrencilere geri dönüt verildi.	5	4	3	2	1
26. Performans görevlerinin önceden planlandığı şekilde hem süreç hem de sonucu notlandırıldı.	5	4	3	2	1
<b>ÖĞRENİM KAZANIMLARI</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Kararsızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
27. Performans görevleri öğrencilerin yazma becerilerine katkı sağladı.	5	4	3	2	1
28. Performans görevleri öğrencilerin okuma becerilerine katkı sağladı.	5	4	3	2	1
29. Performans görevleri öğrencilerin konuşma becerilerine katkı sağladı.	5	4	3	2	1
30. Performans görevleri öğrencilerin dinleme becerilerine katkı sağladı.	5	4	3	2	1
31. Performans görevleri öğrencilerin dilbilgisi yetisine katkı sağladı.	5	4	3	2	1
32. Performans görevleri öğrencilerin kelime bilgilerine katkı sağladı.	5	4	3	2	1
33. Performans görevleri öğrencilerin genel dil yetisine katkı sağladı.	5	4	3	2	1
34. Performans görevleri öğrencilerin otonomilerini (bireysel öğrenme kabiliyeti) olumlu yönde etkiledi.	5	4	3	2	1
35. Öğrencilerin İngilizce öğrenme motivasyonları performans görevleri sayesinde olumlu etkilendi.	5	4	3	2	1

36. Performans görevleri öğrencilerin kendilerine güvenini olumlu etkiledi.	5	4	3	2	1
37. Performans görevleri öğrencilerin dil yetisine katkı sağlamadı.	5	4	3	2	1
38. Performans görevlerinin öğrencilerin derslerde öğrendiklerini pekiştirmeye yardımcı oldu.	5	4	3	2	1
<b>PROGRAMLA UYUM</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Kararsızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
39. Performans görevleri etkili bir ölçme değerlendirme yöntemidir.	5	4	3	2	1
40. Performans görev uygulamaları dil öğrenmede klasik yazılı sınavlardan daha etkilidir.	5	4	3	2	1
41. Uygulanan performans görevleri ders içerikleri ile uyumluydu.	5	4	3	2	1
42. Uygulanan performans görevleri hazırlık programının amaçlarıyla uyumluydu.	5	4	3	2	1
43. Uygulanan performans görevleri performans ölçmeye dayalı sınav prensiplerini yansıtıyordu.	5	4	3	2	1
44. Uygulanan performans görevleri öğrencilerin sınıfta öğrendikleriyle örtüşmedi.	5	4	3	2	1
45. Performans görev uygulamaları klasik yazılı sınavlara sadece destek bir araç olarak faydalıdır.	5	4	3	2	1



**BÖLÜM III.** Lütfen aşağıdaki soruyu hazırlık okulunda ilk dönem uygulanan performansa dayalı ölçme-değerlendirme yöntemi deneyiminiz doğrultusunda yanıtlayınız.

Sizce performans görev uygulamaları gelecek yıllarda devam etmeli midir?

Evet,  ilk dönem uygulandığı haliyle devam etmelidir çünkü

---

---

---

---

Evet,  ama bazı değişiklikler yapıldıktan sonra. Önerilerim:

---

---

---

---

Hayır,  performans görev uygulamaları gelecek yıllarda devam etmemelidir çünkü

---

---

---

---

(Developed by the researcher of the present study)

## **Likert-scale for instructors**

### **CONSENT FORM**

Dear Colleague,

The purpose of this research is to evaluate performance-based assessment as a component of the language program in this EFL preparatory school. This study is being conducted by Nihan ÖZUSLU, master student in the Graduate School of Education at Yeditepe University. I would be pleased if you could help me by responding a questionnaire. As a researcher, I guarantee that all information you provide in the questionnaires will be treated confidentially. The data collected from you is not only vital for my MA study but also for the improvement of the language program in our institution.

Performance-based assessment tasks (PTs), carried out in grammar, vocabulary, reading, and listening-speaking courses, have been administered instead of one pen and paper quiz in 2017-2018 Academic year. This questionnaire, which consists of three parts, aims to gather information on your thoughts about the planning, application, scoring, and learning outcomes phases of PT practices and its consistency with the current language program. Your answers will be used in my MA thesis only by me for academic purposes. If you have any questions about any item in the questionnaire or the research itself, please feel free to ask me.

The participation in this study is voluntary. If you agree to participate, you will spend approximately 15 minutes responding the items. You have the right to withdraw from the study any time you wish. If you wish to be informed about the study results, you may contact me in person or via my e-mail address below.

Thank you in advance for your contribution.

Email: [nihanozuslu@yahoo.com](mailto:nihanozuslu@yahoo.com)

I have read the information above. I agree to be the participant of this research.

Signature:

Date:

**PART I.** This part is related to your personal information. Please put a tick ✓ for your selection.

1. Age: 20-30  31-40  41-50  51 and above

2. Gender: Male  Female

3. Education: Bachelor's  Master's  Doctorate  ELT Certificate (non-ELT BA)

4. Teaching Experience: Less than 5 years  6-10 years  More than 11 years

5. Do you have any experience on designing a test? YES  NO

6. Have you had any testing courses during your education?

YES  Please explain \_\_\_\_\_

NO

7. Have you ever heard about “performance-based assessment” before this term? YES  NO

If yes, in which contexts: Voluntary reading (articles, journals...etc.)  BA courses

MA Courses  MA Thesis  PhD courses  PhD Thesis

Seminars & Conferences  Online teacher forums  Colleagues

Other \_\_\_\_\_



**PART II.** Please display how far you agree or disagree with each of the statements below regarding the PT practices applied in the prep school. Circle the number corresponding to your selection by using the following scale:

5= Strongly Agree    4=Agree    3= Undecided    2= Disagree    1= Strongly Disagree

<b>PLANNING</b>	<b>Strongly Agree</b>	<b>Agree</b>	<b>Undecided</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
1. Sufficient training was provided to me before the PT implementation.	5	4	3	2	1
2. Sufficient training was provided to students before the PT implementation.	5	4	3	2	1
3. Rubrics were explained to me before the PT implementation.	5	4	3	2	1
4. Students did not understand the requirements of the PTs.	5	4	3	2	1
5. Students were given clear instructions about each PT by instructors.	5	4	3	2	1
6. A weekly schedule was announced for each PT.	5	4	3	2	1
7. The deadlines to hand in PTs were announced to the students.	5	4	3	2	1
8. Students were well-informed about the whole process before PT implementation started.	5	4	3	2	1
<b>APPLICATION</b>	<b>Strongly Agree</b>	<b>Agree</b>	<b>Undecided</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
9. PTs were suitable for students' language level.	5	4	3	2	1
10. PTs were suitable for students' cognitive level.	5	4	3	2	1

11. PTs were enjoyable for students.	5	4	3	2	1
12. Sufficient time was given to students for the completion of PTs.	5	4	3	2	1
13. PT implementation was time-consuming for me as an instructor.	5	4	3	2	1
14. Sufficient guidance was provided to me during the PT process.	5	4	3	2	1
15. Students got bored while preparing PTs.	5	4	3	2	1
16. Presenting PTs in class was an effective part of the implementation for students.	5	4	3	2	1
17. Students were able to hand in PTs on time.	5	4	3	2	1
18. Most students prepared their PTs on their own.	5	4	3	2	1
<b>SCORING</b>	<b>Strongly Agree</b>	<b>Agree</b>	<b>Undecided</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
19. Each PT was evaluated through a rubric.	5	4	3	2	1
20. Scoring rubrics were easy to follow for instructors.	5	4	3	2	1
21. I could score PTs objectively.	5	4	3	2	1
22. PT scores were satisfying for most of the students.	5	4	3	2	1
23. Sufficient time was allotted to me for scoring.	5	4	3	2	1
24. I could not score PTs fairly.	5	4	3	2	1
25. I gave feedback to most of the students about their PTs.	5	4	3	2	1

26. I scored both process and product as planned.	5	4	3	2	1
<b>LEARNING OUTCOMES</b>	<b>Strongly Agree</b>	<b>Agree</b>	<b>Undecided</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
27. PTs contributed to the development of students' writing skills.	5	4	3	2	1
28. PTs contributed to the development of students' reading skills.	5	4	3	2	1
29. PTs contributed to the development of students' speaking skills.	5	4	3	2	1
30. PTs contributed to the development of students' listening skills.	5	4	3	2	1
31. PTs contributed to the development of students' grammatical knowledge.	5	4	3	2	1
32. PTs contributed to the development of students' vocabulary knowledge.	5	4	3	2	1
33. PTs contributed to the development of students' general language proficiency.	5	4	3	2	1
34. Students' autonomy seemed to increase through PTs.	5	4	3	2	1
35. Students' motivation to learn English was affected positively through PTs.	5	4	3	2	1
36. PTs had a positive impact on students' self-confidence.	5	4	3	2	1
37. PTs did not contribute to improve students' general language proficiency.	5	4	3	2	1
38. PTs helped students reinforce what they have learnt in the courses.	5	4	3	2	1

<b>PROGRAM CONSISTENCY</b>	<b>Strongly Agree</b>	<b>Agree</b>	<b>Undecided</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
39. Performance-based assessment is an effective assessment method for ELT.	5	4	3	2	1
40. Performance-based assessment methods are more useful than traditional exams for language learning and teaching.	5	4	3	2	1
41. PT implementation applied in this institution was consistent with the content of the courses.	5	4	3	2	1
42. PT implementation in this institution reflected the principles of performance-based assessment.	5	4	3	2	1
43. Performance-based assessment elements were in accordance with the objectives of the language program of this prep-school.	5	4	3	2	1
44. PT implementation conducted in this institution did not reflect what students have learnt.	5	4	3	2	1
45. Performance-based assessment is useful only as a supplementary tool to traditional assessment.	5	4	3	2	1

**Part III.** Please answer the following question based on your own experiences with PTs applied recently in the prep school.

Should PT assessment practices continue in the following years?

•Yes,  in the same way (as applied in fall term) because

---

---

---

---

•Yes,  but only after some revisions. My suggestions are

---

---

---

---

•No,  because

---

---

---

---

(Developed by the researcher of the present study)



## **Likert-scale for students**

### **CONSENT FORM**

Sevgili öğrenci,

Bu araştırmanın amacı performansa dayalı ölçme-değerlendirme yöntemini İngilizce hazırlık okulunda yürütülen öğretim programının bir parçası olarak değerlendirmektir. Araştırma Yeditepe Üniversitesi İngiliz Dili Eğitimi programında yüksek lisans yapmakta olan Nihan Özuslu tarafından yürütülmektedir. Bu anketi yanıtlayarak yapacağınız katkı sadece akademik çalışmam için değil aynı zamanda kurumumuz için de önem arz etmektedir. Fikir ve önerilerinizden öğretim programımızı iyileştirmek amacıyla faydalanılacaktır.

2017-2018 Akademik yılında dilbilgisi, kelime bilgisi, okuma ve dinleme-konuşma derslerinde uygulanan performans görevleri klasik kısa sınav (quiz) yerine geçmektedir. Üç ana bölümden oluşan bu anketin amacı performans görevlerini planlama, uygulama, notlandırma, öğrenim kazanımları ve mevcut öğretim programıyla uyumu açısından sizlerden bilgi toplayarak değerlendirmektir.

Bu anketle elde edilen bilgiler yüksek lisans tez çalışmam için, sadece akademik amaçla kullanılacaktır ve notunuzu etkilemeyecektir. Çalışmadan çıkma hakkınız mevcuttur ve çıkmak istediğinizde bilgileriniz tarafımdan silinecektir. Eğer katılımcı olursanız, yaklaşık 15 dakika ayırmanız gerekmektedir. Çalışma sonuçları ile ilgili bilgi almak isterseniz, benimle aşağıda belirttiğim e-mail adresimden iletişime geçebilirsiniz.

Katkılarınız için şimdiden teşekkürler.

E-mail: [nihanozuslu@yahoo.com](mailto:nihanozuslu@yahoo.com)

Yukarıdaki bilgilendirmeyi okudum. Bu anket çalışmasında katılımcı olmayı kabul ediyorum.

RUMUZ/TAKMA İSİM:

İMZA:

TARİH:

## DEMOGRAFİK BİLGİ

**BÖLÜM I.** Bu bölüm kişisel bilgilerinizi içermektedir. Lütfen sizin için uygun seçeneğe işaret (✓) koyunuz.

- Yaşınız: 18-25  26-30  30 ve üzeri
- Cinsiyet: Erkek  Kadın
- Hazırlık sınıftaki dil seviyeniz nedir? Elementary  Pre-intermediate  Intermediate
- Kaç yıldır İngilizce öğreniyorsunuz? 5 yıldan az  5-10 yıl  10-15 yıl
- Hazırlık sınıfından önce hiç performansa dayalı ölçme değerlendirme yöntemiyle değerlendirildiniz mi?  
EVET  Nerede ve ne şekilde yazınız \_\_\_\_\_  
HAYIR

**BÖLÜM II.** Hazırlık okulunda uygulanan performans görevleri ile ilgili aşağıdaki cümlelere fikren ne derece katılıp katılmadığınızı gösteriniz. Lütfen aşağıdaki sıralamayı kullanarak seçiminize karşılık gelen rakamı işaretleyin:

5= Kesinlikle katılıyorum 4= Katılıyorum 3=Kararsızım 2=Katılmıyorum 1=Kesinlikle katılmıyorum

PLANLAMA	Kesinlikle katılıyorum	Katılıyorum	Kararsızım	Katılmıyorum	Kesinlikle katılmıyorum
1. Öğretmenlerimiz performans görev uygulamaları ile ilgili gerekli bilgiye sahipti.	5	4	3	2	1
2. Performans görev uygulamaları öncesi bizlere gerekli bilgilendirme yapıldı.	5	4	3	2	1
3. Öğretmenlerimiz performans görev sınav yönergeleri (Rubrik) ile ilgili yeterli bilgiye sahipti.	5	4	3	2	1
4. Performans görevlerini nasıl yapacağımı anlamadım.	5	4	3	2	1
5. Performans görev uygulama komutları ve açıklamaları öğretmenler tarafından bizlere net bir şekilde yapıldı.	5	4	3	2	1
6. Performans görevleri için haftalık program önceden duyuruldu.	5	4	3	2	1

7. Performans görev uygulamaları teslim tarihleri bizlere duyuruldu.	5	4	3	2	1
8. Öğrenci olarak performans görev uygulamaları başlamadan önce kendimi bu sürece hazır hissettim.	5	4	3	2	1
<b>UYGULAMA</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Karasızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
9. Performans görevleri dil seviyeme uygundu.	5	4	3	2	1
10. Performans görevleri bilişsel seviyeme uygundu.	5	4	3	2	1
11. Performans görevleri eğlenceliydi.	5	4	3	2	1
12. Performans görevlerini tamamlamak için verilen süre yeterliydi.	5	4	3	2	1
13. Performans görevleri zaman kaybıydı.	5	4	3	2	1
14. Performans görev sürecinde bizlere yeterli rehberlik sağlandı.	5	4	3	2	1
15. Performans görevlerini yaparken sıkıldım.	5	4	3	2	1
16. Performans görevlerini sınıfta sunmak benim için faydalıydı.	5	4	3	2	1
17. Performans görevlerini zamanında teslim edebildim.	5	4	3	2	1
18. Performans görevlerini kendi başıma hazırladım.	5	4	3	2	1
<b>NOTLANDIRMA</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Karasızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
19. Performans görevleri belli bir kritere göre değerlendirildi.	5	4	3	2	1
20. Değerlendirme kriterlerinin anlaşılması kolaydı.	5	4	3	2	1
21. Performans görevleri objektif bir şekilde notlandırıldı.	5	4	3	2	1

22. Performans görev notları tatmin ediciydi.	5	4	3	2	1
23. Ölçme değerlendirmeye ayrılan 1 haftalık süre yeterliydi.	5	4	3	2	1
24. Performans görevlerinin notlandırılması adil değildi.	5	4	3	2	1
25. Performans görevleriyle ilgili bizlere geri dönüt verildi.	5	4	3	2	1
26. Performans görevlerinin önceden planlandığı gibi hem süreç hem de sonucu notlandırıldı.	5	4	3	2	1
<b>ÖĞRENİM KAZANIMLARI</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Karasızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
27. Performans görevleri yazma becerilerime katkı sağladı.	5	4	3	2	1
28. Performans görevleri okuma becerilerime katkı sağladı.	5	4	3	2	1
29. Performans görevleri konuşma becerilerime katkı sağladı.	5	4	3	2	1
30. Performans görevleri dinleme becerilerime katkı sağladı.	5	4	3	2	1
31. Performans görevleri dilbilgisi yetime katkı sağladı.	5	4	3	2	1
32. Performans görevleri kelime bilgime katkı sağladı.	5	4	3	2	1
33. Performans görevleri genel dil yetime katkı sağladı.	5	4	3	2	1
34. Performans görevleri otonomimi (bireysel öğrenme kabiliyeti) arttırdı.	5	4	3	2	1
35. Performans görevleri sayesinde İngilizce öğrenmek için motivasyonum arttı.	5	4	3	2	1

36. Performans görevleri kendime güvenimi arttırdı.	5	4	3	2	1
37. Performans görevlerinin dil yetime katkı sağladığını düşünmüyorum.	5	4	3	2	1
38. Performans görevleriyle derslerde öğrendiklerimi pekiştirdim.	5	4	3	2	1
<b>PROGRAMLA UYUM</b>	<b>Kesinlikle katılıyorum</b>	<b>Katılıyorum</b>	<b>Karasızım</b>	<b>Katılmıyorum</b>	<b>Kesinlikle katılmıyorum</b>
39. Performans görevleri etkili bir ölçme değerlendirme yöntemidir.	5	4	3	2	1
40. Performans görev uygulamaları dil öğrenmede klasik yazılı sınavlardan daha etkilidir.	5	4	3	2	1
41. Uygulanan performans görevleri ders içerikleri ile uyumluydu.	5	4	3	2	1
42. Uygulanan performans görevleri hazırlık programının amaçlarıyla uyumluydu.	5	4	3	2	1
43. Uygulanan performans görevleri performans ölçmeye dayalı sınav prensiplerini yansıtıyordu.	5	4	3	2	1
44. Uygulanan performans görevleri sınıfta öğrendiklerimizle örtüşmedi.	5	4	3	2	1
45. Performans görev uygulamaları klasik yazılı sınavlara sadece destek bir araç olarak faydalıdır.	5	4	3	2	1

**BÖLÜM III.** Lütfen aşağıdaki soruyu hazırlık okulunda ilk dönem uygulanan performansa dayalı ölçme-değerlendirme yöntemi deneyiminiz doğrultusunda yanıtlayınız.

Sizce performans görev uygulamaları gelecek yıllarda devam etmeli midir?

Evet,  ilk dönem uygulandığı haliyle devam etmelidir çünkü

---

---

---

---

Evet,  ama bazı değişiklikler yapıldıktan sonra. Önerilerim: \_\_\_\_\_

---

---

---

---

Hayır,  performans görev uygulamaları gelecek yıllarda devam etmemelidir çünkü

---

---

---

---

(Developed by the researcher of the present study)

## FINDINGS OF OPEN-ENDED SECTION

129 Students:

**Yes, in the same way (Int level 12 sts 30% / Pre-int level 9 sts. 22% /Elementary level 8 sts 17%)**

**Yes, but with revisions (Int level 20 sts. 50% / Pre-int level 22 sts. 54 % / Elementary level 24 sts 52%)**

- Presenting tasks (13%) No videos yes presentations
- Task content (more interesting & creative& funny & authentic tasks) (12%)
- Number of PTs (10%)
- Task materials (8%)
- Less difficult tasks & less demanding (7%)
- More time to complete (6%)
- PT deadlines (5%)

**No, because (Int level 8 sts. 20%/ Pre-int level 10 sts. 24 % Elementary level 16 sts 35%)**

## REVISIONS FOR THE SPRING TERM

**In the light of suggestions from stakeholders:**

- Vocabulary PT timeline was changed (3 PTs –same deadline/ more time to complete PTs)
- Vocabulary task options were lessened
- Reading rubric revision
- Grammar rubric revision
- Grammar through written output
- More pair-work
- Training for instructors (more practical)
- Materials (poster optional)-sending the old ones to the storage
- No CD Yes e-mail of 1 flash disk
- Poster display

Coding: Voc\_int3\_ecesevgi\_task5

(Prepared by the researcher and presented in the PT evaluation meeting)