



**IMPACT OF PREPARATORY SCHOOL’S FINAL ENGLISH
PROFICIENCY EXAM ON STUDENTS’ SPEAKING AND WRITING
SKILLS**

Rukiye Elifnur Özmen

MA THESIS

ENGLISH LANGUAGE TEACHING DEPARTMENT

**GAZİ UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES**

JULY, 2021

TELİF HAKKI VE TEZ FOTOKOPİSİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren altı (6) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Rukiye Elifnur
Soyadı : Özmen
Bölümü : İngiliz Dili Eğitimi
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı: Hazırlık Okulu Yıl Sonu İngilizce Yeterlilik Sınavının Öğrencilerin
Konuşma ve Yazma Becerileri Üzerindeki Etkisi

İngilizce Adı: Impact of Preparatory School's Final English Proficiency Exam on
Students' Speaking and Writing Skills

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazarın Adı Soyadı: Rukiye Elifnur Özmen

İmza :

JÜRİ ONAY SAYFASI

Rukiye Elifnur Özmen tarafından hazırlanan “Impact of Preparatory School’s Final English Proficiency Exam on Students’ Speaking and Writing Skills” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi İngiliz Dili Eğitimi Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Cemal Çakır

.....

Gazi Üniversitesi, Gazi Eğitim Fakültesi Yabancı Diller Eitimi

İngiliz Dili Eğitimi

Başkan: Dr. Öğretim Üyesi Aydan Irgatoğlu

.....

Ankara Hacı Bayram Veli Üniversitesi Yabancı Diller Yüksekokulu

Yabancı Diller Eğitimi

Üye: Prof. Dr. İskender Hakkı Sarıgöz

.....

Gazi Üniversitesi, Gazi Eğitim Fakültesi Yabancı Diller Eitimi

İngiliz Dili Eğitimi

Tez Savunma Tarihi: 02/07/2021

Bu tezin İngiliz Dili Eğitimi Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Yücel GELİŞLİ

Eğitim Bilimleri Enstitüsü Müdürü

.....



Festina lente.
Yavaşça acele et.

ACKNOWLEDGEMENTS

Initially, I would like to thank and express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Cemal ÇAKIR, for his invaluable support during this period. Thanks to his patience, experience, advice, encouragement, and immense knowledge, I was able to write this thesis without difficulty.

Furthermore, I would also like to thank the students and instructors who were eager to take part in this study and share their priceless opinions with me.

Last but not least, my special and sincere thanks are for my family. I would like to thank my mother, Arife ÖZMEN for always being thoughtful, self-sacrificing, my father, Mustafa ÖZMEN for always being understanding and supportive, and my brother, Buğra Nurullah ÖZMEN, his wife, Aylin GÜL ÖZMEN, and my niece Öykü ÖZMEN for their encouragement. I would also love to thank my grandparents and my friends, all of whom have been always by my side and have always believed in me.

**HAZIRLIK OKULU YIL SONU İNGİLİZCE YETERLİLİK
SINAVININ ÖĞRENCİLERİN KONUŞMA VE YAZMA BECERİLERİ
ÜZERİNDEKİ ETKİSİ
(Yüksek Lisans Tezi)**

**Rukiye Elifnur Özmen
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Temmuz, 2021**

ÖZ

Literatürde bir sınavın etkisinin ve geri etkisinin genellikle aynı olduğu düşünülür ve bu kelimeler eş anlamlı olarak kullanılır. Bir sınavın etkisi, geri etkisinden daha kapsamlıdır. Bir sınavın etkisi bir sınavın muhtemel bütün etkilerini kapsar ve sınıf ortamını aşabilir. Bu farklara rağmen etkinin ve geri etkinin derecelendirmesi aynı olabilir: (a) olumlu, (b) olumsuz, (c) hem olumlu hem de olumsuz, (d) ne olumlu ne de olumsuz. Bir sınav öğretmeyi ve öğrenmeyi olumlu bir şekilde etkilediğinde, ‘olumlu etki’ terimi kullanılabilir. Bir sınav öğretmeyi ve öğrenmeyi olumsuz bir şekilde etkilediğinde, ‘olumsuz etki’ terimi kullanılabilir. Eğer bir sınav öğretmeyi ve öğrenmeyi hem olumlu hem olumsuz etkilerse, ‘hem olumlu hem de olumsuz’ terimi kullanılır. Son olarak, ‘ne olumlu ne de olumsuz’ terimi bir sınavın öğretme ve öğrenme üzerinde hiçbir etkisi olmadığı zaman kullanılır. Bunları göz önünde bulundurarak, bu çalışma üniversite hazırlık okulunda yıl sonunda yapılan İngilizce yeterlilik sınavının öğrencilerin konuşma ve yazma becerileri üzerindeki etkisini bulmayı amaçlar. Bu beceriler iletişimsel dil öğreniminde oldukça önemli olmakla beraber değerlendirilmeleri zordur. Bu sınavda öğrencilerin konuşma ve yazma becerileri de test edilmektedir. Çalışmamız, bu durumun öğrencilerin İngilizce öğrenmeye dair duygu ve düşünceleri üzerindeki etkisini de bulmayı amaçlamaktadır. Bu amaçla, Türkiye’deki bir vakıf üniversitesinin hazırlık okulundan 40 öğrenci ve yedi öğretim görevlisi bu çalışmaya katılmıştır. Bu çalışma, nitel bir durum çalışması olarak tasarlanmıştır ve veriler öğrencilerin sözlü ve yazılı ödevleriyle beraber öğrencilere verilen açık uçlu anketten edinilen cevaplardan oluşmaktadır. Bunun yanı sıra, veriler İngilizce öğretim görevlilerinin değerlendirme notları ve yorumlarını da

içermektedir. Bu notlar ve yorumlar değerlendiriciler arası güvenilirliği sağlamak için toplanılmıştır. Öğrencilerin sözlü ve yazılı ödevlerindeki verileri elde etmek için önceden geçerlilik ve güvenilirliği sağlanmış rubrikler kullanılmıştır. Çalışmadaki bütün veriler içerik analizi yöntemiyle incelenmiştir. Elde edilen sonuçlardan yola çıkarak, yıl sonu yeterlilik sınavının etkisinin konuşma ve yazma becerilerinin test edilmesinin etkisinden daha farklı derecelerde olduğunu söylemek mümkündür. Yıl sonu İngilizce yeterlilik sınavının öğrencilerin İngilizce öğrenmeye dair duygu ve düşünceleri üzerindeki etkileri: (a) olumlu, (b) hem olumlu hem de olumsuz, (c) ne olumlu ne de olumsuz, (d) çoğunlukla olumsuz, ve (e) olumsuz. Diğer taraftan, yıl sonu İngilizce yeterlilik sınavında konuşma ve yazma becerilerinin test edilmesinin öğrencilerin İngilizce öğrenmeye dair duygu ve düşünceleri üzerindeki etkileri (a) olumlu, (b) ne olumlu ne de olumsuz ve (c) olumsuzdur. Sınavın içeriği ne kadar genişse etkinin çeşitliliği o kadar çoktur denilebilir. Konuşmanın ve yazmanın test ediliyor olması aynı türde etkilere sahip olmasına ve en baskın etkinin ne olumlu ne de olumsuz etki olmasına rağmen konuşmanın test ediliyor olmasının öğrencilerin İngilizce öğrenmeye dair duygu ve düşüncelerini daha olumsuz etkilediği bulunmuştur. Bunun yanı sıra, yıl sonu İngilizce yeterlilik sınavının öğrencilerin konuşma ve yazma becerileri üzerinde ya olumlu ya da olumsuz etkisi olduğu bulunmuştur. Olumlu etkinin öğrencilerin konuşma ve yazma becerileri üzerindeki etkisinin ne olumlu ne de olumsuz etkiye göre daha güçlü olduğu gözlemlenmiştir. Öğrencilerin fikirlerinin aksine; yıl sonu İngilizce yeterlilik sınavının olumlu etkisi öğrencilerin yazma becerileri üzerindeki etkisine kıyasla, öğrencilerin konuşma becerileri üzerinde daha güçlüdür. Sınavın olumlu etkisinin yeterlilik seviyesi yüksek öğrencilerde konuşma becerileri açısından daha güçlü olduğu gözlemlenmiştir. Bu yüzden, öğrencilerin yeterlilik seviyeleri ve sınavın konuşma becerileri üzerindeki etkisinin gücü arasında bir bağ olduğu söylenebilir. Ancak, bu durum yazma becerileri için geçerli değildir. Öğrencilerin yazma becerileri açısından, öğrencilerin yeterlilik seviyesi ne kadar yüksekse sınavın yazma becerileri üzerindeki etkisinin bir o kadar daha az güçlü olduğu bulunmuştur. Bu yüzden, öğrencilerin yeterlilik seviyeleri ve sınavın öğrencilerin yazma becerileri üzerindeki etkisi arasında bir bağ yoktur. Bu araştırmanın sonuçları öğretmenler, sınav geliştiriciler ve yöneticiler için ikincil değerli bir sınavın önemini anlamalarını ve bu sınavdan yararlı etki elde etme ihtimalini artırmayı sağlayabilecek çıkarımlar yapma fırsatı sunmaktadır.

Anahtar kelimeler :İngilizce öğretiminde sınav hazırlama ve değerlendirme, etkinin çeşitli dereceleri, pozitif etki, negatif etki, geri etki, yazma, konuşma, ikincil değerli sınav

Sayfa Adedi

: 86

Danışman

: Doç. Dr. Cemal ÇAKIR

**IMPACT OF PREPARATORY SCHOOL’S FINAL ENGLISH
PROFICIENCY EXAM ON STUDENTS’ SPEAKING AND WRITING
SKILLS
(M.A Thesis)**

Rukiye Elifnur Özmen

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

July, 2021

ABSTRACT

Washback and impact of an exam are generally considered the same, and washback and impact are used as synonyms in the literature. Impact is more general than washback. Impact involves all possible effects of a test on students, and impact can go beyond the classroom context. Despite these differences, the strength of impact can be the same as the degree of washback: (a) positive, (b) negative, (c) both positive and negative, (d) neither positive nor negative. When the test has beneficial effects on teaching and learning, the term 'positive effect' is used. When a test has harmful effects on teaching and learning, the term 'negative effect' is used. When a test has both beneficial and harmful effects on teaching and learning, the term 'both positive and negative effect' is used. Finally, 'neither positive nor negative' is used once a test has no effects on teaching and learning. By taking these into consideration, our study aims to find the impact of a university preparatory school's final English proficiency exam on students' speaking and writing skills, which are highly important in communicative language learning yet difficult to assess. Furthermore, we aim to find the impact of being tested for speaking and writing skills on students' emotions and opinions. To achieve this goal, 40 students and seven instructors from a foundation university's preparatory school in Turkey took part in the study. The study was designed as a qualitative case study. The data were collected through students' spoken and written assignments and open-ended surveys given to students. The data also included English instructors' assessment notes and comments, which were analyzed for inter-rater reliability. To elicit the data from the spoken and written tasks, the rubrics whose reliability and validity had already been tested were used. All the data were analyzed by applying the content analysis method. Based on the results, it is possible to state that the final English

proficiency exam has more different degrees of impact on students than the impact of being tested for speaking and writing skills does. The impacts of the final English proficiency exam on students' emotions and opinions towards learning English were (a) positive, (b) both positive and negative, (c) neutral, (d) mostly negative, and (e) negative. On the other hand, the impacts of being tested for writing and speaking skills in the final English proficiency exam were (a) positive, (b) neutral, and (c) negative. It can be said that the larger the exam's content is, the more different degrees of impact on the learners the exam has. Although being tested for speaking and writing has the same degrees of impact, and the most dominant impact of being tested for these skills is the neutral impact, being tested for speaking has a more negative impact on students' emotions and opinions towards learning English. It was also found that the final English proficiency exam has either positive or neutral impacts on students' writing and speaking skills. The positive impact was observed more on students' writing and speaking skills than the neutral impact was. In contrast to students' opinions, it was found that the positive impact of the final English proficiency exam is stronger on students' speaking skills than students' writing skills. As for speaking, the higher the students' proficiency level, the stronger the impact is on students' speaking skills. Hence, it might be said that there is a connection between the students' proficiency levels and the degree of impact for their speaking skills. Nonetheless, this cannot be said for students' writing skills. As for writing skills, it was found that the higher the students' proficiency levels are, the less strong the impact on students' writing skills is. Therefore, there is no link between the students' proficiency levels and the degree of impact on their writing skills. The results of this research will have several implications for teachers, test developers, and administrators to see the importance of a low-stake exam and to increase the possibility of yielding positive and beneficial impact from a low-stake exam.

Key Words : Testing in English language teaching, various degrees of impact, positive impact, negative impact, washback effect, writing, speaking, low-stake exams

Page Number : 86

Supervisor : Assoc. Prof. Dr. Cemal ÇAKIR

TABLE OF CONTENTS

TELİF HAKKI VE TEZ FOTOKOPİSİ İZİN FORMU	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
JÜRİ ONAY SAYFASI.....	iii
ACKNOWLEDGEMENTS.....	vi
ÖZ.....	vii
ABSTRACT	ix
TABLE OF CONTENTS	xi
LIST OF TABLES.....	xiv
LIST OF FIGURES	xvi
LIST OF ABBREVIATIONS.....	xvii
CHAPTER 1	1
INTRODUCTION	1
1.1. Introduction	1
1.2. Background of the Study	1
1.3. Statement of the Problem	2
1.4. Significance of the Study.....	3
1.5. Aim of the Study	4
1.6. Research Questions	5
1.7. Assumptions of the Study	5

1.8. Limitations of the Study	6
1.9. Significant Definitions Related to the Study	6
CHAPTER 2	8
LITERATURE REVIEW	8
2.1. Introduction	8
2.2. Language Testing and Assessment	8
2.2.1. Norm-referenced and Criterion-Referenced Tests	10
2.2.2. Proficiency Decisions and Tests	12
2.2.3. A Brief History of Language Testing	14
2.2.4. Reliability and Validity	15
2.3. Washback and Impact	17
2.3.1. Theoretical Framework of Washback	18
2.3.2. Washback Value	20
2.3.3. Related Studies to the Current Research Study	22
2.3.3.1. Washback and Impact Studies Around the World	22
2.3.3.2. Washback Studies from Turkey	32
CHAPTER 3	35
METHODOLOGY	35
3.1. Introduction	35
3.2. Research Design	35
3.3. The Context of the Study	36
3.4. Participants	38
3.5. Data Collection Methods	39
3.5.1. Collection of the Data from Students' Writing and Speaking Tasks	39
3.5.2. Collection of Responses to Open-Ended Survey	45
3.5.3. Review of the Instructors' Assessment Notes and Comments	46

3.6. Methods for Data Analysis	47
CHAPTER 4	48
DATA ANALYSIS & FINDINGS	48
4.1. Introduction	48
4.2. The Analysis of the Students' Tasks and Instructors Assessment Notes and Comments	48
4.3. The Analysis of Open-Ended Surveys.....	50
4.4. Findings	53
4.5. Summary of the Findings	67
CHAPTER 5	69
DISCUSSION & CONCLUSION	69
5.1. Presentation	69
5.2. Discussion.....	69
5.3. Conclusion.....	72
5.4. Suggestions for Further Research	73
5.5. Recommendations for Teachers.....	74
5.6. Recommendations for Testing Units	75
REFERENCES	76
APPENDICES.....	82
Appendix A. Open-ended Survey for Students	83
Appendix B. The Approval of Gazi University's Ethical Comittee.....	85
Appendix C. The Permission of the University Where the Study Took Place	86

LIST OF TABLES

Table 1. <i>Differences between Norm-Referenced and Criterion-Referenced Tests</i>	12
Table 2. <i>Washback Hypotheses</i>	18
Table 3. <i>Some Possible Ways of Promoting Beneficial Washback</i>	21
Table 4. <i>Factors Explaining the Possible Impact of GEPT</i>	29
Table 5. <i>Twelve Elements of Test Factors</i>	30
Table 6. <i>Sato's Hypotheses of the Washback Effect on Learning</i>	31
Table 7. <i>Number of Students and Class Hours</i>	37
Table 8. <i>Tasks for the Speaking Course in the Fall Term</i>	40
Table 9. <i>Tasks for the Writing Course in the Fall Term</i>	41
Table 10. <i>Tasks for the Speaking Course in the Spring Term</i>	42
Table 11. <i>Tasks for the Writing Course in the Spring Term</i>	43
Table 12. <i>The Impact of the End-of-Year Proficiency Exam on Students' Speaking Skills</i>	49
Table 13. <i>The Impact of the End-of-Year Proficiency Exam on Students' Writing Skills</i> ...	50
Table 14. <i>The Impact of the End-of-Year Proficiency Exam on Students</i>	51
Table 15. <i>The Impact of the End-of-Year Proficiency Exam's Speaking Part on Students' Emotions and Opinions towards Learning English</i>	52
Table 16. <i>The Impact of the End-of-Year Proficiency Exam's Writing Part on Students' Emotions and Opinions towards Learning English</i>	52
Table 17. <i>The Impact of the End-of-Year Proficiency Exam on the Improvement of Students' Speaking Skills from Students' Own Perspectives</i>	52

Table 18. <i>The Impact of the End-of-Year Proficiency Exam on the Improvement of Students' Writing Skills from Students' Own Perspectives.....</i>	53
Table 19. <i>Students Who Think Getting Ready for the End-of-Year Proficiency Exam Improved Their Speaking Skills</i>	62
Table 20. <i>Students Who Think Getting Ready for the End-of-Year Proficiency Exam Improved Their Speaking Skills Partially</i>	63
Table 21. <i>Students Who Think Getting Ready for the End-of-Year Proficiency Exam Did Not Improve Their Speaking Skills</i>	63
Table 22. <i>Students Who Think Getting Ready for the End-of-Year Proficiency Exam's Writing Part Improved Their Writing Skills</i>	65
Table 23. <i>Students Who Think Getting Ready for the End-of-Year Proficiency Exam's Writing Part Improved Their Writing Skills Partially</i>	66
Table 24. <i>Students Who Think Getting Ready for the End-of-Year Proficiency Exam's Writing Part Did Not Improve Their Writing Skills.....</i>	67

LIST OF FIGURES

<i>Figure 1. A basic model of washback</i>	<i>19</i>
<i>Figure 2. Traditional washback theory: A stimulus- response model</i>	<i>25</i>
<i>Figure 3. 1990s view of washback: A “black box” model</i>	<i>25</i>
<i>Figure 4. The proposed washback model: A curriculum innovation model</i>	<i>26</i>
<i>Figure 5. Students’ reasons why the exam has positive impact on them</i>	<i>57</i>
<i>Figure 6. Students’ reasons why the exam has negative impact on them</i>	<i>57</i>
<i>Figure 7. Students’ reasons why the exam has both positive and negative impacts on them</i>	<i>58</i>
<i>Figure 8. Students’ reasons why being tested for speaking skills has neither positive nor negative impacts on them.....</i>	<i>58</i>
<i>Figure 9. Students’ reasons why being tested for speaking skills has negative impacts on them.....</i>	<i>59</i>
<i>Figure 10. Students’ reasons why being tested for speaking skills has positive impacts on them.....</i>	<i>59</i>
<i>Figure 11. Students’ reasons why being tested for writing skills has neutral impacts on them.....</i>	<i>60</i>
<i>Figure 12. Students’ reasons why being tested for writing skills has positive impacts on them.....</i>	<i>60</i>
<i>Figure 13. Students’ reasons why being tested for writing skills has negative impacts on them.....</i>	<i>60</i>

LIST OF ABBREVIATIONS

EFL	English as a Foreign Language
CEFR	Common European Framework
TOEFL	Test of English as a Foreign Language
IELTS	International English Language Testing System
ASL	Arabic as a second language
HKCEE	Hong Kong Certificate of Education Examination in English
HKEA	Hong Kong Examinations Authority
EAP	English for academic purposes
GEPT	General English Proficiency Test's
SEPPPO	The Selection Examination for Professional Posts in Public Organizations
ILE	Inter-University Language Examination
IBFLE	Interuniversity Board Foreign Language Examination
SPLE	State Personnel Language Examination

CHAPTER 1

INTRODUCTION

1.1. Introduction

This chapter presents the background of the study, statement of the problem, the purpose of the study, assumptions, limitations of the study, and definitions.

1.2. Background of the Study

Human beings are social creatures. Like food and drink, they need communication to be able to survive, and language is the most common way used for communication. In the earlier times of humans' history, being able to speak the language of the place where you lived was enough to live, study, and work. Nonetheless, this has changed throughout history in parallel with humans' needs. Thanks to the development of science and technology, the boundaries among countries do not have a great effect on people's interactions with each other and communication. English as a Lingua Franca has united almost all of the countries to a great extent. To be able to take part in this global sphere, citizens from different countries had better learn English for various reasons like job opportunities, or academic reasons, and Turkey is no different. In Turkey, students meet English in the elementary second grade at state schools, and their must journey does not end until the final year of high school. If the student decides to continue his/her education as a university student, this journey will probably continue for more years, and it may become a never-ending process for the student. The reason for this is that most of the universities in Turkey offer great percents of courses in English. Moreover, this student needs to be able to be competent in English for good job opportunities.

Being tested is an unquestionable and inevitable part of our daily lives. Naturally, the process of learning and teaching English cannot be thought separately from testing, either.

For an effective educational environment, these three need to go hand in hand, and they always affect each other. This effect might be positive, negative, neutral, beneficial, or harmful (Alderson & Wall, 1993a). As testers and teachers, the intended effect is tests' being beneficial and positive. Nevertheless, things do not always go according to the plan. For example, students studying at the universities which offer courses in English in Turkey must prove that they are proficient enough in English to study the courses in English in the departments. Unless they prove it, they need to study at the preparatory school for a year, and they must prove to be proficient at the end of the year in the final proficiency exam. Needless to say, these exams are important for the test-takers, and they affect learners' learning, teachers' teaching, and testers' testing from different perspectives (Alderson & Wall, 1993a; Hughes, 2002; Shohamy, 2001). Since the general main objective is to enable the students to communicate in English in real life, these tests also try to serve communicative purposes. The more authentic and communicative the tests are, the more possible it is for language teachers and testers to see how the learners are going to act in real life. Nonetheless, communicative testing is not as easy as it sounds (Bachman & Palmer, 1996; Brown, 2004; Carr, 2011). There are many reasons for this such as rater reliability, content validity, time, or money issues. Therefore, the impact of such exams on learners is highly crucial and had better be well understood.

1.3. Statement of the Problem

The limits of my language mean the limits of my world (Wittgenstein, 2010, p. 73). Being proficient in English may give individuals the opportunity not only to eliminate these limits of their own world but also to outshine in different areas like a professional and academic career. Besides, it is mostly a criterion to be able to take place in those areas. Due to these reasons and many others, teaching and learning English have gained importance in Turkey as well. Since the beginning of English language teaching and learning, there have been various improvements with respect to this aim, and still, many more highly developed materials are coming to the stage each day. Some mechanisms naturally appear on the basis of the requirements of the people taking part in this process, and assessment acts as a lead role. When we say assessment, it is a great possibility that testing comes to a person's mind, and language teaching and learning are no different. In Turkey's education system, there are lots of language tests applied to make education

practices standardized. Whereas these tests at the macro level are carried out by governments and prepared by professionals, tertiary tests are present at the micro-level.

Tertiary language testing is quite common in Turkey's education system, and it might not be incorrect to claim that universities comprise an important part of the tertiary education. The universities which offer education in the medium of English, or offer education in English to some degree have their own language testing systems. These tests are sometimes prepared by a permanent testing unit in which the test makers do not teach for longer hours, or do not teach at all, while they are sometimes prepared by teachers who teach fully, and the teacher or teachers who act as test makers always change. According to Shohamy (2001) 's testing development perspective, language testers had better take responsibility both for their tests and for the uses and consequences to which their tests are put. Without being aware of the possible testing consequences and the impacts of these consequences on test-takers' learning process, the intended goal of testing may end up with some serious unintended outcomes.

1.4. Significance of the Study

Speaking and writing are two of the most important communicative skills in English language teaching. Thus, being proficient in these skills is useful both for learners' improvement of their abilities to communicate in real life and for their being successful at tests like the proficiency test. Nevertheless, being proficient in these skills is very problematic for Turkish EFL learners. There are many reasons for this, and not having the opportunity to practise these skills beyond the classroom is one of them. They might do extra and extensive practice to promote their receptive skills (reading and listening) on their own. Nonetheless, it is not quite possible for them to perform their productive skills (speaking and writing), so practising these skills is mostly dependent on the classroom and the target university environment of this study in which these skills are constantly being tested for communicative purposes.

Testing is the most preferred and common way of language assessment, but there are few studies carried out on the subject of testing both in Turkey and other countries, and much rarer studies in the Turkish literature have been performed. The studies in Turkey (Çakıldere & Akpınar, 2013; Dağtan, 2012; Karabulut, 2007; Özmen, 2011a; Özmen, 2011b Sevimli, 2007) have mostly focused on the washback effect of the high stakes tests, which are mainly high school entrance exam or English part of the university entrance

exam or an international high stake exam, or exams made at the tertiary level, and on the perceptions of academics of some national high stake exams. These show us that the impact and washback effect of low stakes test have not been investigated substantially despite their common usage. The reason for this might be the idea that these low stake tests do not produce washback, or they do not affect teaching and learning. On the other hand, this belief may not be based on a plausible ground. The reason for this is that these tests are mostly regarded as important by testers, test-takers, and administrators, and they are constantly given to testees. . According to Shohamy (1998), the growing impact of quite many classroom tests might substantiate the final grade, which could yield an eventual high stake decision. Based on these grounds, a study carried out by Köktürk (2015) focused on the washback effects of examinations at the university level on teaching and learning of English. The study carried out was result-driven, and excluded the washback effect of the end of year proficiency examination due to the time issue.

Based on Alderson and Wall's washback hypotheses' (1993a), it would not be wrong to state that a test will affect learning from different angles. In parallel with this idea, Cheng and Curtis stated "there are always consequences" for test-takers (2004, p. 89). These outcomes might be both process-oriented and results-driven, and most of the research studies performed in Turkish and foreign literature are result-driven excluding the impact of the process. Hence, there is a gap in both pieces of literature with regard to the impact of a test which focuses on the process rather than the outcome. The present research was carried out to narrow this gap in the literature. Because of the stated reasons, this study is important, and it will contribute to the English language testing literature

1.5. Aim of the Study

The university in which the study took place is a foundation university in Turkey. Due to the use of the medium of English, students need to study a preparatory school to be proficient in English before they study at their own departments. After studying one term, students might take a proficiency exam, and they might continue studying in their departments so long as they pass. If not, they need to continue their study at the preparatory school for one more term, and they have to take the end of the year proficiency exam. However, taking this proficiency exam, which is administered at the end of the first term, and passing it is not quite common. This proficiency exam is a type of communicative testing. It aims to test four skills of listening, reading, writing, and

speaking. Therefore, this present study aimed to find the impact of this exam on students' productive skills (speaking and writing), which are highly important in communicative language learning yet difficult to assess.

1.6. Research Questions

In line with this purpose given above, answers to the following questions were sought.

1. Is the end-of-year English proficiency exam similar to high stake international communicative tests in terms of its structure, reliability, and validity?
2. Is the end-of-year English proficiency exam communicative enough?
3. How does the end-of-year proficiency exam affect the learners? Does this exam have any positive, negative, or both positive and negative effects on learners?
4. What do students think about the impact of being tested for the speaking skill in the end-of-year English proficiency exam on their opinions and emotions towards learning English?
5. What do students think about the impact of being tested for the writing skill in the end-of-year English proficiency exam on their opinions and emotions towards learning English?
6. How does the end-of-year English proficiency exam affect their productive skill of speaking?
7. How does the final English proficiency exam affect their productive skill of writing?

1.7. Assumptions of the Study

1. The participants would be sincere, eager, and honest while answering the interview questions.
2. The participants would have enough amount of knowledge about the proficiency exam and its requirements.
3. The participants would do the given speaking and writing assignments meticulously and honestly.

1.8. Limitations of the Study

This is a qualitative case study. Although the results were quantified, the data were collected at only one foundation university's preparatory school via an open-ended survey, assignments of the students, assessment rubrics, and instructors' notes and comments. The research context is only limited to one universe. Therefore, the results cannot be generalized, which limits the external validity. Another limitation of this study is the time limit. Shohamy, Donitsa-Schmidt, and Ferman (1996) state that washback possibly evolves over time. Although this is not a result-driven study, it would be better to see the proficiency exam's results of the participants and to create a link between their learning process and their exam results. Another limitation is the COVID-19 pandemic. Due to the pandemic, the lessons were covered online. Last but not least, the validity and reliability of the proficiency test were also examined by the researcher; however, it might not be reflective of the real examination process enough, since the researcher was not able to observe the examination procedure.

1.9. Significant Definitions Related to the Study

Assessment: Getting information about students, educational policy, and curricula to be used during decision-making process (Nitko & Brookhart, 2013).

Norm-referenced tests: Test-takers' performances are interpreted with a comparison to how other test-takers' have done. The scope of testing is general, and it is not based on a specific criterion. According to their results, test-takers are placed on a continuum, and a normal distribution is expected based on these results (Brown, 1996; Brown, 2004; Carr, 2011).

Criterion-referenced tests: Test-takers' results are independently interpreted on the basis of well-defined criteria or objectives. A test-taker's performance is not judged in a reference to other test-takers' performance. Based on their results, test-takers also get feedback about how they have performed (Brown, 1996; Brown, 2004; Carr, 2011).

Foreign language proficiency: The ability to speak, write, read, or listen in a non-native language. There are different proficiency levels. According to Common European Framework (CEFR), there are six proficiency levels varying from beginner (A1) to native-like (C2).

Proficiency tests: The test types administered with the aim of predicting the test-takers' future language use. The language backgrounds of the test-takers are little or no important (Hughes, 2002).

Validity: Validity in language testing means that the test should measure what it intends to measure (Brown, 2004; Carr, 2011; Fulcher & Davidson, 2007). Roughly, if the test aims to measure the ability of making an oral complaint about waiting for long hours, then the test had better make the test-takers complain about waiting for long hours during the test.

Reliability: Reliability can be defined as consistency of scoring (Bachman, 1990; Brown, 2004; Carr, 2011; Green, 2013). When the same test is applied to the same test-takers in the different time span, the test-takers' scores should be the same or similar to their earlier scores for the sake of reliability of the test.

Impact: Impact includes all possible effects of a test possibly have "on individuals, policies or practices, within the classroom, the school, the educational system, or society as a whole" (Wall, 1997, p. 291).

Washback effect: The effects that the tests have on teaching and learning (Alderson & Wall, 1993a; Gates, 1995; Hamps- Lyons, 1997; Hughes, 1989; Messick, 1996; Tsagari, 2007; Wall, 2012; Watanabe, 2004).

Positive (beneficial) washback effect: An intended or a beneficial effect of a particular test teaching and learning (Alderson & Wall, 1993a).

Negative (harmful) washback effect: An unintended or a harmful effect of a particular test teaching and learning (Alderson & Wall, 1993a).

Strong washback effect: Strong washback effect takes place when a particular test changes teachers' and students' behaviours to a great extent with the aim of being successful (Alderson & Wall, 1993a).

Weak washback effect: Weak washback effect occurs if a particular test has little or no washback effect on teachers' and students' behaviors (Alderson & Wall, 1993a).

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

In this chapter, the related literature to the present study is explained. Firstly, literature related to testing is provided, focusing on testing and assessment, a brief history of testing, and test types. Next, the washback effect is explained in terms of washback effect (positive, negative, neither or both), how washback works, washback and validity, and theoretical framework of washback. Finally, washback and impact studies around the world and in Turkey are summarised.

2.2. Language Testing and Assessment

There is a general misconception about testing and assessment being synonyms. Tests are administered during a specific time constraint, and the learners are aware of the fact that their answers will be scored and evaluated. On the other hand, assessment is an ongoing process. Moreover, there are a number of ways of assessments, and testing is just one of them. Everything which happens inside the classroom now and earlier, or assignments which are done outside the classroom could be assessment tools. These include formal or informal student observations, how a student has performed on an assignment, portfolios, and analyses of students' earlier records (Nitko & Brookhart, 2013). The stated misbelief above may stem from the fact that tests are the most common assessment tools in most areas including language testing and assessment. Regardless of being so popular and present in almost every phase of our lives, feeling unpleasant is not surprising even when we are used to being tested, and we happen to be so confident in the area in which the test is going to be. However, this is not always unavoidable. Brown, maintains that tests “ need

not be degrading, artificial, anxiety provoking experiences” (2004, p. 2). The secret of achieving this goal is lying beneath the way how the tests are designed and administered.

So what is a test? According to Brown (2004) “A test, in simple terms, is a *method of measuring a person's ability, knowledge, or performance in a given domain.*” (p. 3). In this definition, some terms were pointed out and exemplified in a detailed way by Brown (2004). Therefore, it will be best to look at them specifically to have a better understanding of this definition. First of all, an effective test should have a method which is supposed to be straightforward and shaped. The second important asset is measuring. Unless a test presents a result to the test-takers, it is of no use to administer it. Next, who the test-takers are should be known by the testers. No matter how effective and professional the test is, it is meaningless if it does not address the needs of the test-takers. Bachman and Palmer (1996, pp. 4-6) offered an exquisite example with this regard. They initially had the idea that they would end up with an effective way of testing when they modeled their own tests to large scale EFL tests. Nevertheless, they came up with miscellaneous unintended outcomes from different perspectives like participants’ backgrounds and goals. Fourthly, it is true that a test measures performance, but at the same time it had better give a clue about the test takers’ ability; that is, their competence. Last but not least, the test should measure in the given domain. For instance, the test should only measure the vocabulary items taught in a specific unit if it is meant to measure the vocabulary items taught in the specific unit. Another definition of a test was stated by Bachman. According to Bachman, a test is “a measurement instrument designed to elicit a specific sample of an individual’s behavior.” (1990, p. 20), so it can be said that a test is different from measurement in that it is prepared for getting a specific type of behaviour.

In terms of testing, Sadler (1983) stated that testing is complementary for effective teaching and a crucial part of effective teaching, so the significance of language tests is beyond doubt. Initially, they could be precious instruments to supply information relevant to various concerns in language teaching and learning. Besides, they can put forward feedback about how effective the education system is since they can yield proof of learning and instruction’s results. By taking the learners’ strengths and weaknesses enlightened by language tests into consideration, teachers will have a detailed idea about the learners’ needs, and they will give learners more effective and on-point materials. Eventually, language tests help the illumination of instructional objectives, and this may help the teacher to have a clearer idea of what to teach. Regardless of being so important, effective

language tests are difficult to prepare and administer. Hence, being competent in language testing bears great importance. Unfortunately, there may be such misbeliefs and misconceptions about language tests as not understanding the nature tests and not having a plausible expectation (Bachman & Palmer, 1996, p. 20), which may make the testers have unintended results unless they are competent enough in language testing.

Unequivocally, language tests are vital, but too much importance has been on the evaluation function of the tests by ignoring their instructional functions. After the focus was shifted to the instructional function of the tests, the washback term came to the stage. In the coming sections of this chapter, washback, which has led this research to be conducted, will be explained in a detailed way.

2.2.1. Norm-referenced and Criterion-Referenced Tests

The testing scope of norm-referenced tests is more general than that of criterion-referenced tests (Brown,1996; Brown, 2004; Carr, 2011). The best and common example of norm-referenced tests could be The Test of English as a Foreign Language (TOEFL). Even though TOEFL has subtests, these subtests are quite general. On the other hand, criterion-referenced tests, as the name suggests, are based on clearly stated objectives or criteria. Along with the results, the test-takers also get some feedback from their teachers about their performance. Classroom-tests used at English medium universities, which are associated with a curriculum and objectives focusing on what students can or cannot do, might be the perfect example for criterion-referenced testing. Another key difference between the two tests is the interpretation of scoring (Brown,1996; Brown, 2004; Carr, 2011). By looking at the score of a norm-referenced test, it is not possible to judge how well the test taker performed in the test. It can only be known how successful the test taker was according to the other test-takers. Let us have an imaginary test taker called X. When X scored 70 percent in a criterion-referenced test, it would mean that X answered 70% of the questions correctly in the test, and other test-takers' performances do not affect X at all. On the contrary, when X scored 70 percentile in a norm-referenced test, it would mean that X's score was better than 70 students' scores out of 100 students. There is no way to know whether X's performance in the test was good or not since X's performance is judged with a reference to other test takers' performances. It can be said that criterion-referenced tests are more prone to yielding positive washback effect. In that, test-takers are able to see their progress thanks to criterion-referenced tests , which are based on clear objectives and

programmed to give feedback to the learners. The only thing that matters here is to pass or fail without being affected by what others have done. In norm-referenced tests, our test-taker X can have a desirable place in the continuum when s/he can do better than the others. However, this does not necessarily mean that X meets the standards nor has s/he performed effectively enough. Briefly, test-takers only compete with themselves in criterion-referenced tests, while they compete with others regardless of their own being effective in norm-referenced tests. That is why criterion referenced tests could possibly promote more beneficial washback effect than norm-referenced tests do (Hughes, 2002). In norm-reference tests, a normal distribution is expected, and the scores are often presented in percentiles (Brown,1996; Brown, 2004; Carr, 2011; Nitko & Brookhart, 2013). Another vital difference is the purpose of these tests. Judging from the interpretation ways, the aim in norm-referenced tests is comparative, and the test-takers' scores are presented in a continuum, while the scores in criterion-referenced tests are absolute. They assess how much the test taker knows, or what the test taker is able to do. The other distinctions based on the various characteristics are categorized in Table 1 below.

Table 1

Differences between Norm-Referenced and Criterion-Referenced Tests

Characteristic	Norm-referenced	Criterion-referenced
How it is interpreted	The test-takers' performance is illuminated with a comparison to other test-takers'.	The test-takers' performance is explained with a comparison to the amount how much s/he has learned.
How it is measured	General language abilities and proficiencies are measured.	Specified objectives are measured.
Why it is used for testing	To lay out the test takers' along a continuum of general abilities	To assess how much the test-takers' know or have learned the material
How the scores are distributed	A normal distribution is expected around the mean.	Various scores (If the test easy, the scores are higher. If the test is difficult, the scores are lower.)
What the test structure is like	Consists of moderately long subtests including diverse question contents (Since the question scope is general, it is difficult to guess the questions. Only the questions are known by the test-takers.)	Short clarified questions sets with similar contents (The questions are based on clear objectives. Therefore, the questions are more guessable for the test takers.)

Source: Adapted from Brown, J. D. (1996). *Testing in language programs*. New Jersey: Prentice Hall Regents.

In conclusion, it is difficult to make a decision about whether a test is criterion-referenced or norm-referenced just by looking at it. Similarly, it is not easy to decide which one is to be used as the measuring instrument in one's testing context. The milestone during this decision- making process should be the well-clarified purposes. With the help of these clarified purposes, language teachers will possibly choose the best test type along with the measuring tool for their learners.

2.2.2. Proficiency Decisions and Tests

Proficiency tests are used in a great diversity of fields and areas to see if a person is proficient enough to carry out certain skills or jobs. Language testing is one of those areas. For instance, when a company looks for a person with the ability of writing 100 hundred

words in a minute, a proficiency test could be applied. Similarly, English language proficiency tests at English medium universities are also administered both at the entrance and exit to see if an individual is proficient enough to carry out his/her study and career requiring effective use of English. First of all, the question of what proficient means needs to be answered. One definition of 'proficient' is being well advanced in an art, occupation, or branch of knowledge (merriam-webstercom) In terms of language testing, being proficient means having enough requirements of the language for a particular purpose (Hughes, 2002). Proficiency decisions in language testing are to be taken into consideration when the external world is supposed to be related to a program. Despite not being associated with a certain syllabus, program, or curriculum, proficiency tests' specifications which determine whether the test-taker is competent or not should be stated clearly and in a detailed way (Brown, 2004; Brown, 1996; Carr, 2011; Heaton, 1990; Hughes, 2003). Besides how the language constructs are stated is the primary factor in proficiency tests. These tests should be prepared by professional testers due to its general nature (Brown, 2004; Brown, 1996; Carr, 2011; Hughes, 2002). Therefore, it would be better for language teachers not to create their own proficiency tests.

In nature, proficiency tests are general, and they entail the test-takers' general level of language competency by comparing each test-taker with other test-takers. These two characteristics of proficiency tests make the interpretation of the results of these tests norm-referenced (Brown, 2004; Brown, 1996; Carr, 2011; Hughes, 2002). These facts might hinder the possibility of beneficial washback; however, proficiency tests are the most functional ones (regarding time and money issues) to answer some questions. Some of these questions are whether the students have sufficient knowledge of English to follow a certain programme at university, or whether the students are capable of using English for professional purposes when they leave these programs (Brown, 1996; Heaton, 1990). Also, proficiency tests are necessary for sustaining the unity of present institutions and keeping individuals from studying the program which they indeed do not need to study.

The vital point here is being careful while deciding proficiency decisions and choosing these tests carefully. Also, the possibility that each institution might be in need of having a different proficiency test should not be ignored during a decision-making process.

2.2.3. A Brief History of Language Testing

To have a better understanding of classroom-based testing and communicative language testing, it is better to have a look at the brief history of language testing (Bachman, 1990; Brown, 2004; Brown, 1996; Carr, 2011). Along with the shifts in language teaching methodology, the language testing was also shaped. Grammar-translation approaches led the prescientific movement to appear in language testing. During this movement, teachers were free to create their own tests mainly consisting of free compositions and translation tests. This naturally put subjectivity in danger. Following the prescientific movement, psychometric-structuralist movement and integrative-sociolinguistic movement came to stage respectively. The former one was based on behavioral psychology and structuralist linguistics. During this movement, discrete-point tests were popular. Discrete-point testing supported the idea that language components should be tested by getting broken into smaller units (Brown, 2004; Carr, 2011), and decontextualization was favoured. Although this movement increased the concerns related to the objectivity of scoring, reliability, and validity, it caused the authenticity to be problematic. The latter one (integrative-sociolinguistic movement) considered language components to be tested as a whole unlike the discrete-point testing based on the psychometric-structuralist movement. Despite the scoring procedure being more difficult than the discrete-point testing, the integrative testing was more real-life like; therefore, the communicative language use was more necessary (Carr, 2011). Test types during integrative-sociolinguistic movement were dominantly cloze-tests and dictation. Cloze tests consist of a short reading passage (150-300 words) with some deleted parts, requiring the test-takers to fill in those blank parts, and dictation makes the test-takers write what they hear correctly by listening to the a passage an administer reads aloud (100-150 words).

Later on, the communicative approach sometimes being linked to the integrative approach was an initiative for communicative movement. Unlike integrative approach, the communicative approach is concerned with how language is used for communicative purposes (Heaton, 1990). Bachman and Palmer's (1996) opinion about language tests is that test performance should represent the language use not only in the testing environment but also in the non-testing environment in order to provide useful and intended outcomes. This idea also points out the necessity of communicative language testing. The earlier contexts of language testing were not natural and spontaneous. Therefore, they were failing to mirror the language use in real life situations (Brown, 2004; Carr, 2011). Along with the

communicative language testing taking its roots from Canale and Swain's (1980), and Bachman's (1990) communicative competence model, authenticity and genuineness have also gained importance. Naturally, the importance of culture in language testing has gone up. The test-taker might simply fail to speak or write about a content when the content is far from being relevant to his/her culture. For instance, a test prepared by British people and administered in Japan resulted in unintended outcomes for oral part (Heaton, 1990). In that part, test-takers were required to complain about a person smoking; however, complaining is not a part of Japanese culture especially about this kind of trivial complaining. This cultural bias certainly affected the reliability of the test. Therefore, test administrators should be careful about the content and its relevance to the culture for the sake of reliability.

In terms of the results of communicative language testing, various profiles of test-takers could be taken. We can take the imaginary test-taker X (mentioned above in the norm-referenced and criterion-referenced part) into consideration. If X took a communicative proficiency test, X's performance of four skills would be measured. Along with this, X's results might differ. In that, X might get a high score in the reading part, while s/he could get a low score in the writing part. This stems from communicative language test's miscellaneous measuring of proficiency more than as just one total score. In parallel with this, communicative language testing introduced qualitative type of assessment which shows the test-takers their different language performance levels via language band systems. Of course, this puts the objectivity of scoring at stake to a certain degree, but reliability is increased by using clear specifications and criteria.

To sum up, tests had better be based on "clear definitions of both the abilities we wish to measure and the means by which we observe and measure these abilities" (Bachman, 1990, p. 81) as long as intended purposes are targeted.

2.2.4. Reliability and Validity

When reliability and validity come to people's minds, the focus is generally on the differences between them. However, they may be better understood when they are seen as supplementary for each other. The aim of reliability is to decrease the errors stemming from measurement, and the purpose of validity is to keep the effect of the language abilities we target to measure at the maximum level (Bachman, 1990).

Reliability roughly can be defined as the consistency of scoring (Bachman, 1990; Brown, 2004; Carr, 2011; Green, 2013). Some categories of reliability are test reliability, test-administration reliability, student reliability, and rater reliability. The first one is about the nature of a test. To increase the test reliability, the test should be short, clear, and well-written. The second one is about the setting in which the testing takes place. For instance, the audio should be heard well enough by every test-taker if the test includes the listening part. The third one is about how the test-takers are feeling. The illness or fatigue could be considered as factors which affect this reliability. Finally, rater reliability can be categorized as intra-rater and inter-rater reliability. The former one occurs when the same rater has inconsistent results in different times, while the latter one happens when two or more different raters have inconsistent results.

Validity in language testing means that the test should measure what it intends to measure (Brown, 2004; Carr, 2011; Fulcher & Davidson, 2007). For example, if the ability to write a formal e-mail is aimed to be tested, it should be tested in the test by making the test-takers write a formal e-mail during the test. In the testing literature, there are five main types of validity. These are content validity, criterion-related validity, construct validity, consequential validity, and face validity (Brown, 2004; Carr, 2011; Fulcher & Davidson, 2007). Content validity is related to how much the test represents the whole universe. It is about how well the test reflects the curriculum or the subject matter on which the test is based (Carr, 2011). Criterion-related validity is interested in finding out whether the criterion of the test has been achieved or not. This is divided into two types. They are concurrent validity and predictive validity. The first one is carried out with a comparison of two different assessment types, and the other one is the link between the current scores and future success. Before addressing what construct validity is, let us define the construct. First of all, construct ought to be explained “in such a way that it becomes measurable” (Fulcher & Davidson, 2007). Basically, construct validity is the theory behind the test goal; therefore, it would not be wrong to say that it is the milestone of testing. Consequential validity, as the name suggests, covers all the possible consequences of a test like intended, unintended social consequences, or test preparation process. Washback and impact of a test may be considered under the title of consequential validity. Face validity, last but not least, is about how the test is seen by the test-takers. To increase this validity, the test should be familiar, doable in the time constraint, or well-constructed.

As a summary, validity and reliability are the two most important facets of language testing. Due to the paradigm shifts in language teaching, the importance of authenticity and impact which are directly related to validity has also been pointed out (Davidson & Cumming, 2007). In order to sustain an effective language teaching, learning, and testing environment, these two aspects should be well comprehended and applied along without ignoring the authenticity and impact factors.

2.3. Washback and Impact

Even though impact is sometimes used as a synonym for washback, impact is more general than washback. Impact includes all possible effects of a test possibly have “on individuals, policies or practices, within the classroom, the school, the educational system, or society as a whole” (Wall, 1997, p. 291). Unlike washback, tests’ impact could reach beyond the classroom walls and could affect society. Especially for the high-stakes tests, this is neither a surprising nor an unusual fact in the Turkish context. The impact of test use works in two ways. One of them is the macro level which is related to the matters of the test use in the scope of the educational system or society. The other one is micro level which deals with the impact of test use in the view from participants affected by the test (Bachman, 1990; Bachman & Palmer, 1996).

Washback is one of the several dimensions of tests’ impacts (Bachman & Palmer, 1996; Tsagari, 2007). Regardless of a growing interest in washback, one will not possibly have a clue about what washback is just by looking up the dictionary. The reason for this is that the definition of washback is not present in the dictionaries. However, several definitions could be found for backwash, which is used widely and interchangeably for washback, in some dictionaries. A dictionary defines backwash as “ the unpleasant after-effects of an event” (lexico.com), and this definition surprisingly has a negative connotation. In the testing literature, there are various definitions of washback. Nevertheless, washback is commonly defined as the effects that the tests have on teaching and learning (Alderson & Wall, 1993a; Gates, 1995; Hamps- Lyons, 1997; Hughes, 1989; Messick, 1996; Tsagari, 2007; Wall, 2012; Watanabe, 2004). Since being used synonymously with backwash whose definition has negative connotations, washback’s first impression may be negative as well. However, this is not the case. The term washback is neither positive nor negative, but it is related to the impact of the test. If the test is not effective enough, the washback effect might be felt negative or vice versa (Alderson & Wall, 1993a). The washback effect

is a complex issue and dependent on a number of variables. Likewise, yielding a positive, negative, neither or both washback effect is also dependent on those variables. This issue shall be discussed in the further chapter.

2.3.1. Theoretical Framework of Washback

One important issue related to washback in the literature is that there has not been enough research carried out in order to understand the very nature of washback and how its mechanism works. It would not be wrong to state that Alderson and Wall (1993a) could be considered as a milestone for washback studies. After they had brought up the question of whether washback exists or not in 1993a, the influences of tests got more attention for research.

Alderson and Wall recommended future researchers to try to specify The Washback Hypothesis by including classroom observations into their research and embodying findings in the fields of motivation and performance along with the educational innovation. Based on the literature review and research they had done, Alderson and Wall came up with some possible washback hypotheses which could shed light for washback researchers (Table 2).

Table 2

Washback Hypotheses

-
1. *A test will influence teaching.*
 2. *A test will influence learning.*
 3. *A test will influence what teachers teach.*
 4. *A test will influence how teachers teach.*
 5. *A test will influence what learners learn.*
 6. *A test will influence how learners learn.*
 7. *A test will influence the rate and sequence of teaching.*
 8. *A test will influence the rate and sequence of learning.*
 9. *A test will influence the degree and depth of teaching.*
 10. *A test will influence the degree and depth of learning.*
 11. *A test will influence attitudes to the content, method, etc. of teaching and learning.*
 12. *Tests that have important consequences will have washback.*
 13. *Tests that do not have important consequences will have no washback.*
 14. *Tests will have washback on all learners and teachers.*
 15. *Tests will have washback effects for some learners and some teachers, but not for others.*
-

Source: Alderson, I. C., & Walt, D. (1993a). Does washback exist? *Applied Linguistics*, 14, pp. 120-121.

The washback effect on participants, teaching, and learning can easily be seen on the table.

Another view about how washback works was stated by Hughes (1993). He took washback into consideration from three perspectives. These are participants, process, and product. Regarding this view, the test affects the participants' perceptions and actions towards teaching and learning. As a result, the participants' performance could get affected, which refers to the process. Finally, the product based on the quality of learning, and what is learned appears on the stage (Hughes, 1993). Based on these two views (Alderson & Wall, 1993a; Hughes, 1993), Bailey (1996) put forward a model for washback (Figure 1).

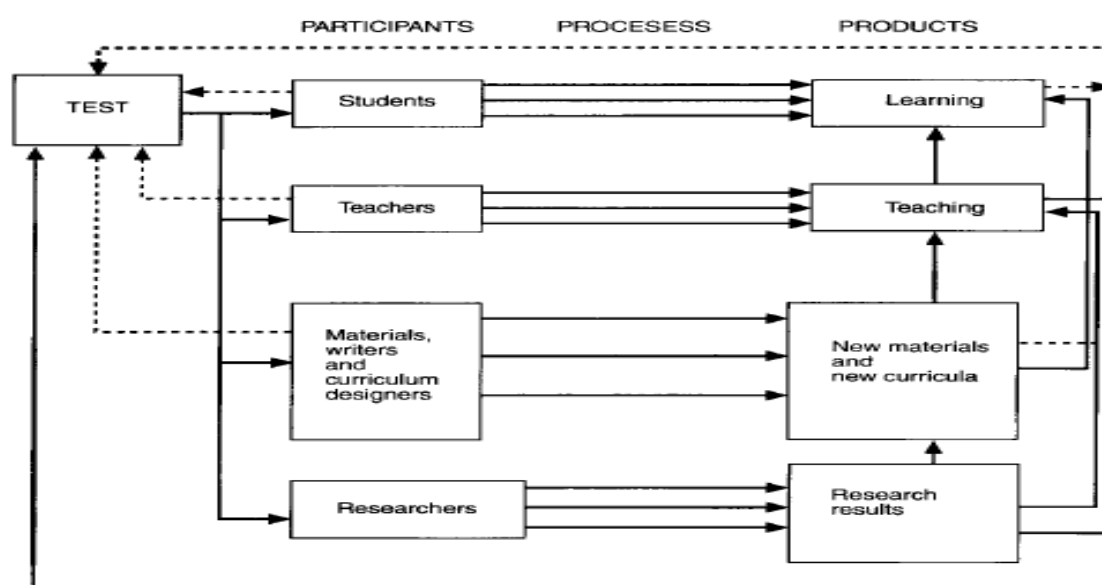


Figure 1. A basic model of washback (Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), p. 264)

Bailey examined washback under two categories. These are washback to the learner and washback to the program. Washback to the learner enlightens how the learners act when they face an important test, and it is directly related to Alderson and Wall's washback hypotheses number 2,5, 6, 8, and 10 (Bailey, 1996). Briefly, these processes include learners' study habits, perceptions, and attitudes affected by the test. When we consider the Turkish context, learners consistently object to being tested, and the field of language teaching and learning is not different. Besides, these tests are mostly high-stakes tests, and they are life-changing occasions for the learners. Hence, washback to the learner should be well understood for the sake of a high quality testing environment. In the latter one, i.e., washback to the program, the tests' impact on teachers, test administers, and curriculum developers are discussed, and it is linked with Alderson and Wall's washback hypothesis number 1, 3, 4, 7, 9, and 11 (Bailey, 1996). Based on the test results, important

decisions are made in terms of participants, educational system, or other parties in charge. Therefore, it is not easy to take responsibility of not teaching for the test (Qi, 2005). Here, washback to the program should be taken into consideration cautiously since the danger of a narrowed curriculum is highly possible based on the washback effect on the programme.

2.3.2. Washback Value

In terms of language testing, the washback effect is generally regarded as negative at first. Nonetheless, the washback effect can be strong, weak, positive, negative, beneficial, harmful, none or all (Alderson & Wall, 1993a; Bailey, 1996; Cheng & Curtis, 2004; Gates, 1995; Hughes, 2002; Messick, 1996). When the tests' effect encourages teachers and learners to change their behaviors to get high marks in the test, the effect will be strong. On the contrary, the tests' effect will be at the minimum level or none at all when the test has a weak effect. Gates (1995) listed some factors which affect the strength of washback: "prestige, accuracy, transparency, utility, monopoly, anxiety, familiarity, and practicality" (pp. 102-103). Prestige refers to the fact that the test will have a stronger effect if it is popular like International English Language Testing System (IELTS) or Test of English as a Foreign Language (TOEFL). Based on this, it is possible to claim that standardized-tests will have stronger effects. Accuracy is about the reliability of the scores. Since the language test results are used for life-changing occasions, accuracy and reliability are important. Transparency is related to a test being direct or indirect. Regarding this, the more direct the language test is, the more transparent it is. Hence, the possibility of a strong washback effect is higher. Utility refers to the functionality of the test. It is about the number of the doors a test could open based on the results taken. Utility and monopoly factors are indeed related to each other. If a test has fewer competitors in the language field, the washback effect will be stronger. Even though anxiety may sometimes be a negative and deterring factor in language teaching, learning, and testing, a plausible amount of it might promote the strength of the washback effect. Finally, practicality, as the name suggests, is about how appropriate a test is in terms of time and money issues.

Alderson and Wall (1993a) explained negative washback as an unintended effect of a particular test, which is considered 'poor', on teaching and learning. They further stated that The Washback Hypothesis is seen to presume that "teachers and learners do things *they would not necessarily otherwise do* because of the test." (p. 117). Here, it is possible to infer that a poor test might bring about positive outcomes as long as it encourages

positive attitudes and behaviors. Let us take a poor classroom test into consideration. No matter how poorly it is designed, the effect could still be positive when it makes the learners prepare for the tests properly, or do their assignments more eagerly. Negative and harmful washback effects also occur when the non-tested skills or items are ignored (Green, 2013; Hughes, 2002). For example, the English part of the university entrance exam in the Turkish context does not test productive skills (speaking and writing). As a result of this, test-takers can be claimed not to practise these skills. This is directly related to construct under-representation which occurs when the test fails to include important dimensions of the main constructs. Likewise, the negative and harmful effect will be the possible outcome once the test evaluates the skills or items which are not in the curriculum (Green, 2013). This time, construct-irrelevant variance which happens when the test includes too much reliable variance not being related to the interpreted construct is at stake.

The positive washback effect is believed to take place once the course objectives and test concur (Gates, 1995; Green, 2013; Messick, 1996). According to Bailey (1996), promoting beneficial or positive washback should be the preliminary purpose for test administrators. Hughes (2002) and Bailey (1996) came up with some pieces of advice with this regard (Table 3).

Table 3

Some Possible Ways of Promoting Beneficial Washback

1. Testing the abilities whose improvement we want to stimulate 2. Sampling broadly and unpredictably 3. Testing directly and in an authentic way 4. Making testing criterion-referenced 5. Basing achievement tests on clear objectives 6. Tests' familiarity and understanding for teachers and students 7. Learner autonomy and self-assessment 8. Providing assistance to teachers if necessary 9. Counting the cost
--

Sources: Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), pp. 268-272; Hughes, A. (2002). *Testing for language teachers*. Cambridge: Cambridge University, pp. 53-56.

This table could be helpful to increase beneficial washback. Nonetheless, it had better be kept in mind that these recommendations are mostly for classroom achievement tests. Besides, the quality of washback is not only dependent on the quality of the test but also dependent on different variables like how the teachers teach, and what educational system

or the assessment setting is (Alderson & Wall, 1993a; Messick, 1996). In the language testing literature, there is not a consensus about if the washback effect is positive or negative (Cheng & Curtis, 2004). For a more broad consideration of tests' influences, it is better to carry out impact studies (Hamp-Lyons, 1997).

2.3.3. Related Studies to the Current Research Study

To have a look at the washback and impact studies carried out so far would be useful in terms of understanding the value and mechanism of washback and impact. Therefore, these related studies are going to be presented in this chapter as two categories. These are washback and impact studies around the world and washback and impact studies in Turkey.

2.3.3.1. Washback and Impact Studies Around the World

Alderson and Wall (1993b) carried out an impact study in Sri Lanka, where English is a second language, and students study English there from grade 3 to grade 11. At the end of grade 11, students take an O-level exam which includes English and other curriculum subjects. In this exam, the weight of English is important either for students' academic career or for job opportunities. The purpose of this study was to find out whether O-level examination had an impact on the process of English Language Teaching in the Sri Lankan secondary schools. This study combined classroom observations with traditional forms of data collection such as questionnaires, interviews, and test analyses. The observation programme was carried out by 7 teachers in 5 different parts of Sri Lanka in the years 1990 and 1991. The observation process was divided into 6 rounds in the two academic years. Round 1 was considered as a pilot round. Although rounds 2, 3, and 4 supplied useful information for the further rounds, their scopes were not large enough, and some of them were affected by political issues. Because of these reasons, observations of rounds 5 and 6 were largely focused on. Round 5 took place 5 months before the exam with 61 classes, and round 6 occurred 1 month before the exam with 41 classes. Based on the findings, the number of classes that use the textbook during class time decreased in round 6. Similarly, the focus on the tested skills went up whereas the focus on the not tested skills went down. In conclusion, the findings showed us that there were both positive and negative washback effects on the content of teaching. As for positive washback effect, covering the textbook could be judged as a positive effect unless it was only done like memorization for the

exam. For negative washback effect, narrowing the curriculum was stated. In that, the tasks related to tested skills were pointed out, while the tasks related to not tested skills were less emphasized. In terms of methodology, there was no evidence for washback. Alderson and Wall (1993b) also stated that teachers did not have sufficient idea about what and how the test measures. Like the content of teaching, there was both evidence for positive and negative washback effects regarding the ways of assessing. There was more notice in reading and writing than grammar, which was positive. The item-types were similar to the ones in the textbook, which was positive, too. Nonetheless, washback effect on the item-types based on the textbook were negative if they were used too much. This study was significant from different perspectives. One of them is that the data collection was triangulated by using different tools. Moreover, the study shows washback researchers the significance of answering the question of what washback is. As an extension of this study, Wall (2005) carried out another investigation based on the same data as Alderson and Wall's (1993b). The motivation behind this investigation was the questions coming from the original project, and the investigation was interested in both the findings and the context in which the earlier study was carried out. Thanks to this study, she found the fact that the intended washback effect was not only dependent on the test itself but also it was dependent on features of the educational setting, the learners, and teachers. Therefore, the intended effect could be different based on these factors. It could not be so wrong to express that this study provided the people to whom it may concern with some recommendations about the possibility of a test to influence the changes in the curriculum.

Shohamy, Donitsa- Schmidt, and Ferman (1996) performed a washback study which examined the impact of two national tests in Israel. These were the effects of an Arabic as a second language (ASL) examination for grades 7-9, and an English as a Foreign Language (EFL) oral test for grade 12. The former one is a low-stake test, while the latter one is a high-stake test. They collected the data from students through questionnaires consisting of Likert-scale items and open-ended questions. The data from teachers and inspectors were collected via interviews. Finally, analysis of the Director-General Bulletins (1984-96) and instructions issued by the Ministry of Education were performed. The results indicated that the impact of the ASL test had decreased over time. Teachers stated that they neither did sufficient preparation for the exam nor they informed their students about the exam. In parallel with teachers' statements, it was found that students and their parents were hardly aware of the presence of the test. Unlike the authorities' beliefs, the

test's prestige among both students and teachers was found pretty low. Teachers' viewed this exam as an ordinary occurrence without any consequential effects and poor in quality. In terms of the EFL oral test, it was found that there was a significant increase in the impact of the test regardless of how small the modifications were. Both students and teachers regarded this exam as important, and they stated that they cared about the matter of success and failure in this exam to a great extent. Shohamy et al. (1996) added some concluding remarks in terms of these huge differences among ASL and EFL tests. These differences were mainly ascribed to the characteristics of these tests like being high or low-stake, language status, the purpose of the tests, the format of the tests, and the skills tested in each test.

Cheng's washback (1997 & 1999) studies were both carried out in Hong Kong upon a change in a high-stakes public examination which is the Hong Kong Certificate of Education Examination in English (HKCEE). The change was administered by the Hong Kong Examinations Authority (HKEA). The former study (1997) aimed to find out whether the change had any washback effects on the teaching of English at the secondary schools, while the latter one's (1999) purpose was to investigate how this change may affect classroom teaching "in relation to the reactions of the local education context and also with regard to aspects of teachers' attitudes and behaviours" (Cheng, 1999, p. 257). In both studies, observations had a crucial role in terms of data collection tools, and both of them include interviews and surveys as well. Both studies' findings were almost similar. According to the findings of the first study, the change in the examination caused an intense washback effect in the teaching content. Some washback effects on teachers' behaviours and attitudes were also observed. Nonetheless, the collected data were not large enough to judge whether this effect was positive or negative. The findings of the second study revealed the appearance of new activities in the classrooms and a great change in the teaching content. As a result of the encouragement of HKCEE and teachers' understanding of this encouragement, teachers used miscellaneous activities in their classrooms regardless of not having enough knowledge of the new exam format. However, the change in the teaching approach of the teachers and in the interaction pattern between teachers and students was observed to be very little. This is a concurrent finding with Alderson and Wall's study (1993b) which is that a change in the exam might affect what to teach but not how to teach.

Burrows (1998; 2004) conducted a research study in Australia to find out the differences between teachers' classroom practices before and after the introduction of a new competency based curriculum and assessment by considering teachers' belief systems and their resultant response to the change. She implemented surveys, interviews and observations into her data collection procedure. While she was collecting the data, some flaws occurred regarding the instruments. However, this led the data collection procedure to be more beneficial. Based on the findings coming from interviews, it was stated that respondents experienced changes which stem from either the implementation of assessment or the implementation of curriculum. Once the respondents were asked to state the changes resulting from implementation of assessment and curriculum, their answers were more or less the same. Therefore, the differences between them were not significant. The reason for this outcome might be "the "connectedness" of the assessment and the curriculum" (Burrows, 2004, p. 119). Another important result of the interviews showed that the amount of change which occurred to the teachers were not the same. The observations in the classes were videotaped and analyzed later. The findings of the observations led the researcher to coming up with a proposed washback model (Figure 4). The observations showed that the choice factor, from the teachers' perspectives, was involved in washback effect to a certain degree at least in classroom-based assessment context.

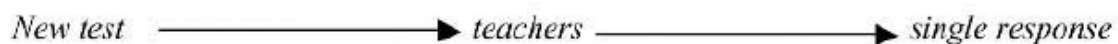


Figure 2. Traditional washback theory: A stimulus- response model (Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Research Contexts and Methods*, p. 126)

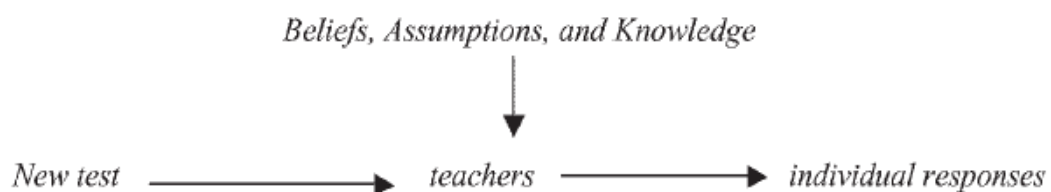


Figure 3. 1990s view of washback: A "black box" model (Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Research Contexts and Methods*, p. 126)

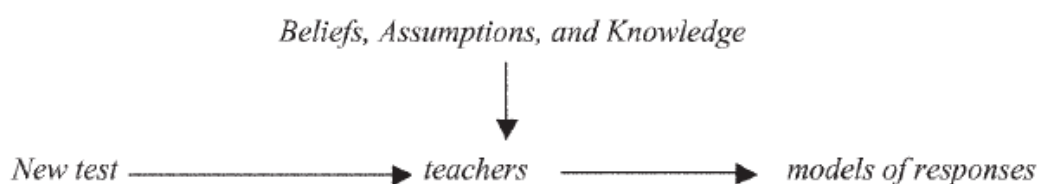


Figure 4. The proposed washback model: A curriculum innovation model (Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Research Contexts and Methods*, p. 126)

The traditional model comes before Alderson and Wall's washback model, and this washback model does not take teachers into consideration as a variable for the of washback. The "black box" model favours objective data gathering. The results are based on emprical studies. The proposed washback model, on the other hand, stands for the idea that "that a qualitative analysis of teachers' responses to the introduction of a new test or assessment system may reveal patterns in their responses" (Burrows, 2004, p. 126). As concluding remarks, this study showed that there might seem to be a connection between curriculum and washback. Besides, the data collection procedure proved the fact that things not going according to the plan would not always pose a risk for the study. On the contrary, they could yield much better results than previously expected.

Watanabe (2004) investigated the teacher-related factors affecting the washback through interviews and detailed classroom observations which took part in the school setting. Here the studied test was the university entrance exam. Before the procedure of the data collection, the researcher predicted that both positive and negative wahsback effects would be present in the research context. The predicted negative washback effects could be tagged as "undesirable" classroom teaching activities like too much dependence on grammar-translation method, the persistent reference to exams, and more focus on the formal structures of English rather than real-life like communicative situations (Watanabe, 2004, p. 133). On the other hand, the only positive predicted washback was students' being motivated by the exams. The findings of the study were in parallel with the hypothesized negative washback effects; however, this was not the case for all the teachers who took part in the study. In that, the observed frequency of the negative washback effect was not the same for all teachers, and the researcher stated that the reasons behind this effect might not only stem from exams. Teachers' beliefs, backgrounds, and bias could be considered as factors resulting in these effects, too. The researcher also stated that there were cases which

contradicted the predictions and could be explained as an indication for the existence of positive washback. As concluding remarks, firstly, it was presented that only some of the reported negative washback effects of the entrance exam in the mass media were present, but other ones were not. Secondly, the causes explaining what teachers were doing resulted from various factors and not just necessarily stemmed from the exam itself. Thirdly, teachers' psychological factors appeared to be involved in intervening in the process of washback. For example, the bias against the exam could be a weakening factor for teachers, while being familiar with different teaching methods could be supportive for them. Last but not least, the school context may also be a mediating factor. For instance, one of the school contexts which was a part of the research study held different activities like international concerts. According to the researcher, this positive and colourful school atmosphere might influence the exam classes in a beneficial way, which may end up with engineering the positive washback effect. To sum up, judging from the study's findings, it is possible to say that entrance exam brought about some negative outcomes to some extent for teachers' lessons and some possible washback effect as long as the teachers used the exam preparation as an opportunity to improve their students' proficiency levels. This study once more showed that the washback effect was not only dependent upon the test's validity and reliability. Promoting beneficial washback lies beneath solving the problems related not only to the individual and school but also to the society.

Hayes and Read's (2004) focused on the impact of IELTS test on international students' preparation in New Zealand. In terms of triangulation, different data collection tools like classroom observations, teacher interviews, and teacher and student questionnaires were used. The results focused on especially two schools which took part in the study, and these schools were tagged as School A and B. The findings showed a clear positive washback effect of IELTS preparation at School A. However, this washback effect was not as positive as expected in that, the focus was more on the practice of test tasks rather than the general academic proficiency. It may not be wrong to infer that the test impact somehow ended up with a narrowing in the curriculum. On the other hand, it was found that the course seemed to point out a broader scope of academic study and to encourage students' general language development. Since the length of these course and students background differed, the mentioned conclusions could be deceptive.

Green (2006) conducted a research study in order to investigate the possible impact of IELTS and in order to find out how instruction differed between IELTS preparation and,

English for academic purposes (EAP) classes in academic writing. The study took place in the United Kingdom. As for the data collection, different instruments like observations and interviews were used, and various universities, colleges, and language schools took part in this study. This study focused on the differences between IELTS preparation and EAP classes. Nevertheless, just like the washback model hypothesized, significant areas such as organisation, content control, and student modality overlapped according to the findings. On the other hand, variations in the language content were not significant. The researcher pointed out the fact that the differences between classes (IELTS and EAP) might also stem from teachers' backgrounds, training, and their beliefs about effective learning. Moreover, interviews with teachers showed that their approaches toward the two classes were not the same in spite of their content being similar.

Most of the studies in the literature, including the ones mentioned above, more focused on the washback effect on teaching and teachers. Unlike them, Shih's washback study (2007) was carried out to investigate General English Proficiency Test's (GEPT) on learning in Taiwan. The focus of his study was merely on learning and learners, and it investigated the effect of a proficiency exam. Since the current study also aims to investigate the washback effect of a proficiency exam, Shih's study is rather more significant to be mentioned. As research contexts, the researcher chose two different universities. In one of them (University A), GEPT is not a must; however, day-division students have to pass the first part of GEPT's intermediate level or the make-up exam administered by the school at the other one (University B). Researcher collected the data through interviews, observations, and reviews of his own research notes. The results showed that GEPT had a small impact for the participants at University A. At University B, the results were not far from University A. In that, GEPT had a slightly higher degree of washback among University B's students than A's. Another finding presented the fact that students, relatively, prepared more for reading and listening skills which were tested in the first stage. On the other hand, speaking and writing skills were primarily tested in the second stage. When compared with the preparation of receptive skills (listening & reading), preparation of productive skills (speaking & writing) is much more demanding and requires other people involved. This showed that the nature of the tested skill is as important as the test structure. The researcher listed 9 factors (Table 4) which account for the possible impact of GEPT.

Table 4

Factors Explaining the Possible Impact of GEPT

1. The GEPT did not have any immediate importance.
2. English was the student's major.
3. Students were not aware of how to prepare for GEPT or had nowhere to practise.
4. Students' proficiency was higher than the required to pass GEPT.
5. Students were not available before the test.
6. Students' self-described laziness prevented their test preparations from happening.
7. Students' GEPT preparations mirrored their learning attitudes.
8. Students did not support the notion of learning for the test.
9. Make-up examination as means of escape

Source: Adapted from Shih, C. M., (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), pp. 144-150.

Considering these, the researcher concluded that there were extrinsic and intrinsic factors which affect the impact and washback along with the test factors. Due to the inability to explain these changes, the researcher came up with a provisional washback model. In this washback model, main categories were extrinsic, intrinsic, and test factors. These main categories also had their own sub-categories. Socio-economic factors; school and educational factors; family, friends, and colleagues; and personal factors were brought together under the category of extrinsic factors. On the other hand, intrinsic factors include the ones related to the individual himself/herself. These are individual differences, personalities, and individual perceptions of the test. The test factors have 12 elements some of which were concurrent with the previous studies like Shohamy's study (1996), and some of which were suggested by the researcher himself (Table 5).

Table 5

Twelve Elements of Test Factors

1. The test's stake affects the washback effect.
2. The immediate importance of the test must be considered.
3. A language test is more possibly to generate an impact on learning once the tested skills are rarely practised by the learners.
4. The difficulty of the tests influences learners' test preparation process.
5. The loopholes related to the test will probably lessen the impact of the test.
6. The test affects the learning content of the learners.
7. The effect of the test structure should not be ignored.
8. The nature of the tested skills influences the impact of the test.
9. Socio-political status of the language has a partial impact on the degree of washback.
10. How the test is administered affects the level of anxiety.
11. The test-takers' preference could be more prone to being tested for more than one skill.
12. The purpose of the test may change the consequences. Hence, it might affect the degree of washback.

Source: Adapted from Shih, C. M., (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), pp. 152-153.

As a conclusion, this model pointed out the fact that tests would affect the learners' future learning. Therefore, the tests should be designed by taking this into consideration. In spite of the time constraint and the limited universe of this research, and the researcher's recommendations bear a great importance for the washback literature.

Muñoz and Álvarez's study (2010) was aimed to specify the washback effect of an oral assessment system on certain areas of EFL teaching and learning. The researchers collected the data by using teacher and student surveys, classroom observations, and external evaluation. In the research study, the participants were from two separate groups which were experimental and comparison groups. The researchers had some principles used as a guide for the investigation of washback so as to examine the research questions. These principles were divided into two categories which were teaching and learning (Muñoz & Álvarez, 2010, pp. 5-6). Teaching included a match between curriculum objectives, and instructional tasks, various assessment tasks and task design, and well-clarified and specific feedback. On the other hand, learning included comprehension of self-assessment criteria and learner self-assessment. The results, discussion, and conclusions were presented based on these guidelines. They observed a degree of positive washback effect among the various features of teaching and learning. The results showed that continuous guidance and support to teachers and students increased awareness of the

assessment procedures. Likewise, clarification of objectives, rubrics, and structured assessment tasks would possibly increase the beneficial washback effect. Moreover, it was found out that the application of self-assessment mechanisms promoted beneficial washback effect in that, learners who apply self-assessment mechanism will probably become more aware of and responsible for their own learning, and thus; more focus-oriented and self-directed learners. Lastly, this study suggested precious information for curricular decisions.

In spite of its limitations among which the lack of triangulation for the qualitative data is a serious one, Sato's study (2019) is worth mentioning. Sato's study investigated components affecting Japanese learners' English learning during their preparation for the university examination. The researcher collected his data through interviews. He found out that students' learning behaviours were affected by exams, their own views, school, and exam-independent factors such as their hobbies and interests. Based on the results and findings, he came up with some hypotheses (Table 6).

Table 6

Sato's Hypotheses of the Washback Effect on Learning

-
1. The more common the test items are in the different types of exams, the more students' learning is influenced if the students are preparing for more than one exam.
 2. Grammatical and lexical knowledge are considered as basic knowledge, skills, and abilities (KSAs) for the reading comprehension by the students. Therefore, they are still studied even when grammar and vocabulary items are not directly tested.
 3. Students decide their weaknesses by taking their perceptions of test tasks' difficulty into consideration, and they have a clue about which KSAs are required for them based on these weaknesses. They try to deal with these weaknesses by engaging in various activities like revising their past exam papers.
 4. Classroom activities and advice from teachers have an effect on students' learning behaviour; however, this effect is shaped by students' views on learning methods, as well. Unless students think that this is an effective method for their learning, they do not apply them. Their own learning experiences form the perceived effectiveness of learning methods.
-

Source: Adapted from Sato, T. (2019). An investigation of factors involved in Japanese students' English learning behavior during test preparation. *Papers in Language Testing and Assessment*, 8(1), p. 90.

Sato stated that the above mentioned occurrences had not been presented by earlier studies. Hence, these hypotheses might be added to the existing theories of washback on learning if

they are empirically tested. These hypotheses could be a starting point for the future researchers who want to explore the washback effect on learning.

Hung and Huang's research study (2019) investigated washback effects of a proficiency test –in a Taiwanese campus-wide English curriculum context and their relationships with learner characteristics. The researchers collected the data via questionnaires. The collected data were not supported neither by interviews nor by observations. Therefore, like Sato's study (2019), the findings from the questionnaire lack validation coming from qualitative data. The findings of the study were concurrent with the earlier washback studies which proposed that washback effects on learning were present. The results showed that washback effects of the test were especially notable on such areas as personal image, future job opportunities, motivation and emotion of learners whereas there was less washback effect observed in terms of the interaction between teachers and peers, which was contradictory to the findings of the earlier studies. According to Shih's model of washback (2007), it is possible to tell that extrinsic aspects of a test's washback on learning were weaker than the intrinsic aspects. The results also showed that there was a significant link between the strength of washback and students' proficiency levels in that, the degree of washback was stronger for high-proficiency level students. On the other hand, no significant difference was observed in terms of the connection between learner characteristics, which were gender and academic background, and the strength of the washback effect. To sum up, this study showed that providing learners with a sustained help and guidance might yield positive washback, and this was parallel with the suggestion made by Muñoz and Álvarez (2010) that students could focus more on their learning thus eventually be more successful so long as they had enough knowledge of the assessment practices (Hung & Huang, 2019, p. 14). Although it is difficult to generalize the findings of the study to the whole universe, they can be useful for teachers, policy makers, and school administrators, as well. They may focus on increasing the strength of washback effect on low-proficiency level students, and they can ignore the effect of gender and academic background to some extent during the decision-making process.

2.3.3.2. Washback Studies from Turkey

To find out the washback effect of the Selection Examination for Professional Posts in Public Organizations (SEPPPO) in Turkey, Özmen (2011a) carried out a study with participants who wanted to be English teachers. Although the participants took this exam

with the aim of being English teachers, their major departments varied from ELT to linguistics, and English or American literature. The researcher collected the data through questionnaires and semi-structured interviews. The study found that SEPPPO yielded negative and harmful washback effects. These findings were categorized as some themes, like educational and social factors. For the educational effects, the exam was not considered as trustworthy or valid by the participants. In terms of the social effects, SEPPPO was seen as a burden not only for test-takers but also for their families and education programs. Student-teacher participants stated that they had not had enough time to meet the requirements of the courses and teaching practicum due to preparing for SEPPPO. Here, it can be inferred that negative or harmful washback effects have an effect on different parties such as “social groups, individuals and institutions” (Özmen, 2011a, p. 43). Another washback study carried out by Özmen (2011b) investigated the washback effect of Inter-University Language Examination (ILE) on candidate academics taking this exam for various academic goals. The data collection tool consisted of only interviews. The results of this study presented that ILE had a negative impact on candidate academics, and this negative impact was mostly because of the test’s content and style. Besides, this negative impact was observed both at the micro level including individuals or small groups, and at the macro level like a larger group of people in university contexts. A year later, Dağtan (2012) carried out a similar study on the washback effect of ILE. Nevertheless, the researcher translated this exam’s name as Interuniversity Board Foreign Language Examination and abbreviated it as IBFLE. Although Dağtan (2012) analyzed different aspects of this exam from Özmen’s study (2011b), there were some concurrent areas like the emphasized skills, the effect of IBFLE on test-takers’ preparation, and the required changes in the content of the exam. In terms of the content of the exam, Dağtan (2012) stated that the participants did not think of this exam as the right tool to assess language skills, and it should have tested other main skills like listening, speaking, and writing like TOEFL did. This finding was in line with Özmen’s (2011b) finding. In contrast to Özmen’s (2011b) findings on the test-takers’ preparation for this exam and the emphasized skills, Dağtan (2012) found out that the test-takers attached importance to non-tested skills such as speaking and writing, too, during their preparation for the exam.

Duran (2011) carried out a washback study of a low-stakes exam. This study was carried out mainly to see the teachers’ and students’ perceptions of the washback effect of classroom-based speaking tests. Furthermore, students’ and teachers’ beliefs about testing

and teaching speaking were addressed as well. The study was carried out in a state university's preparatory school located in the southern part of Turkey. 45 instructors and 307 students took part in this study voluntarily, and the data were collected via questionnaires and interviews. Duran's (2011) study revealed that both students and teachers have positive attitudes towards learning/teaching and testing speaking skills. In terms of the washback effect of this classroom-based speaking test, no washback effect on what teachers teach or learners learn was observed. However, both students and teachers stated that they would continue to learn/teach speaking skills even when they were not tested.

Çakıldere and Akpınar's study (2013) was based on two high-stakes language exams which were administered nationally in Turkey. These were State Personnel Language Examination (SPLE), and Inter-University Language Examination (ILE). The study aimed at the impact of these two high-stakes exams on receptive and productive skills of academic personnel from a state university in Turkey. For data collection, questionnaires and surveys were used. Based on the results, the researchers stated that SPLE and ILE had a positive washback effect on reading which was the only skill tested in these exams. On the other hand, untested skills of speaking, listening, and writing were completely ignored by the participants. The results of this study were parallel with those of Karabulut (2007) and Sevimli (2007), both of which investigated the washback effect of the university entrance examination's English language part.

Köktürk (2015) aimed to find the washback effect of low-stakes exams on teaching and learning. These exams investigated in the study consist of quizzes and mid-terms. The study was carried out in a state university's preparatory school located in the central parts of Turkey with both student and teacher participants. The data collection instruments were both questionnaires and interviews. The results indicated that the quizzes and midterms had both positive and negative washback effects on teaching and learning. When the quizzes' washback effects were compared to the midterms' washback effects, it was identified that midterms yielded more positive washback effects than quizzes. This might have stemmed from the fact that midterm exams tested four skills. Another finding showed that the midterms had more positive washback effects on productive skills than receptive skills. The reason may be about the resemblance between the content taught in the classroom and the content tested in the test.

CHAPTER 3

METHODOLOGY

3.1. Introduction

In this chapter, research design and data collection procedures are presented. First, the research design is explained, then the context of the study and participants are given. Finally, data collection instruments and data analysis methods are provided.

3.2. Research Design

This current study follows a case study method in a qualitative design. The quantitative data were only used in order to present some parts of the results. In the literature, a case study is often defined as an in-depth examination of an occurrence that is unique and bounded to a system in its natural context (Adelman, Jenkins, & Kemmis, 1976; Cohen, Manion, & Marrison, 2000; Gall, Borg, & Gall, 2003; Hitchcock & Hughes, 1995; Merriam, 1998; Stake, 2000; Yin, 2008). There are different types of case studies described in the literature. Since our case study is more akin to one of Stake's types of case studies, a brief explanation of the Stake's is provided. Stake (2000) presented three types of case study: intrinsic, instrumental, and collective. In the intrinsic case study, comprehending a specific case is the preliminary goal, and the case itself is the main point. As the name suggests, in the instrumental case study, the case itself acts as an instrument which gives a supportive insight into a theory or an issue. The case itself here is the second concern. Last but not least, the collective case study consists of multiple cases in order to explain an event. By its nature, this current study is also an instrumental study. Although each preparatory school in Turkey follows a different program, almost all of them have an end-of-year proficiency exam. Therefore, the findings of this current study could be supportive for them, too. Gall, Gall, and Borg (2014) stated four main characteristics of a

case study: its focus is on a certain occurrence of an event; it requires a deep examination of the event; it explores the event in its natural context; it stands not only for the participants' viewpoints (emic) but also for the researchers' viewpoints (etic). This current study is a case study since it focuses on an event which occurs in a bounded system, bounded time (7 months) and place (one preparatory school of a foundation university). It also provides a deep understanding of the event (the impact of end-of-year proficiency exam on students' speaking and writing skills) in its natural context. Besides, this study has both the researcher's and the participants' viewpoints. Another reason for naming this study as the case study is that the case study is to be applied when a researcher wants to shed light on the question of "how" (Merriam, 1998; Yin, 2008), and the researcher wants to find answers for questions of "how". Merriam (1998) presented three main features of a qualitative case study. These are particularistic (the focus is on a specific event or case), descriptive (the final product of a case study is rich thanks to the thick description, and heuristic (a case study enlightens the reader's comprehension of the studied event). In this case study, the generalization of the results to other contexts or cases will not be possible; however, some recommendations could be supplied for the other contexts. Nonetheless, this negative aspect could be overcome with a detailed explanation of the case (thick description), like Merriam (1998) presented, and using different data collection instruments (triangulation).

3.3. The Context of the Study

The study was conducted at a preparatory school of a foundation university in the Eastern-Mediterranean part of Turkey. For the academic year 2020-2021, there were 447 students enrolled in this preparatory school program. The students who will study their department courses fully or to some extent in English take a proficiency exam at the beginning of each academic year. This proficiency exam consists of grammar, vocabulary, listening, reading, and writing parts. The students who pass the exam directly start studying in their departments, and the ones who fail take a placement test. This school lets the students take an end-of-term proficiency exam. If they pass, they will continue their studies in the departments. However, the students generally do not prefer to take this exam, and the ones who take it rarely pass. The placement test consists of grammar, vocabulary, listening, reading, and writing parts. According to the results of the placement test, the students are placed into their classes to study at the preparatory school. There are three different

language level types of classes at this school: beginner (A1), elementary (A2), and pre-intermediate (B1). For the Academic Year of 2020-2021, there are 23 beginner groups and 279 beginner students. Except three groups which have 20 students, the others have 11 students each at the beginner level. The students at this level take 28 hours of English courses a week, 20 hours of which are for the coursebook. In the coursebook courses, the focus is on grammar and vocabulary. Since grammar is taught in the context, they practice the other four skills as well. Four hours are for each of the reading-writing and listening-speaking courses. For the elementary level, there are five groups and 106 students. Four groups have 21 students each, and one group has 22 students. At the elementary level, the students take 26 hours of English courses a week, 16 hours of which are for the coursebook. Five hours are for the reading and writing courses, and the other five hours are for the listening and speaking courses. Finally, there are three groups and 62 students at the pre-intermediate level. The two of these groups have 21 students, and one of them has 20 students. These students take 24 hours of English courses a week, 12 hours of which are for the coursebook and 6 hours are for each of listening-speaking and reading-writing courses. All of these class hours and the number of students are summarized in Table 7 below.

Table 7

Number of Students and Class Hours

Level	Number of Students	Coursebook hours	Listening & Speaking Course Hours	Reading & Writing Course Hours	Total number of class hours
Beginner (A1)	279	20	4	4	28
Elementary (A2)	106	16	5	5	26
Pre-intermediate (B1)	62	12	6	6	24

Because of COVID-19, all of these courses were covered online. Each classroom has three different instructors, coursebook, reading-writing, and listening-speaking. While all of the listening-speaking instructors and most of the reading-writing instructors are foreigners, coursebook instructors are Turkish.

For the assessment and testing in the 2020-2021 Academic Year, the number of exams and quizzes that each level takes is the same. There are three types of testing in this context: quizzes, progress tests, and the end-of-year proficiency exam. The progress tests measure

students' grammar and vocabulary knowledge; and listening, reading, writing skills. Progress test is applied to see the students' development. In all levels, the students take one coursebook quiz. For the listening-speaking and reading-writing courses, they take two quizzes each course. For the exams, they take two progress tests in each level. Besides the exams and quizzes, the students are also assessed by tasks for their productive skills (speaking & writing). There are five tasks each for all levels for the writing and speaking courses each semester. As a summary, students are responsible for 10 quizzes, 4 progress tests, 20 tasks, 1 end-of-year proficiency test during the Academic Year of 2020-2021. To be able to pass the preparatory school, each student's mean score should be minimum 50. The grades they take from these tasks, quizzes, and progress tests affect %40 of their final score. The end-of-year proficiency test affects %60 of their final score. However, each students has to get more than 50 from the end-of-year proficiency exam for their results to be calculated.

For both speaking and writing courses, all levels use the book entitled *Unlock (2nd Edition)*. Beginner level students use Book 1, elementary level students Book 2, and, finally, pre-intermediate level students Book 3 in the fall term.

As for the fall and spring term, students are responsible for five speaking tasks, and they are required to hand in five writing tasks each term. The explanations about fall and spring terms and the dates of these tasks are provided in the data collection methods part.

Students from all levels continue using the *Unlock (2nd Edition)* book in the spring term, but the level of the books each level uses changes. Beginner level students use Book 2, elementary level students Book 3, and, finally, pre-intermediate level students Book 4.

3.4. Participants

A case study can be carried out in only one site or multiple sites; however, the characteristics of the sites should be similar to end up with mutual themes (Cresswell, 2013). Since the preparatory schools in Turkey follow different programs, it is difficult to do this study in different sites. Therefore, the current study has focused on one site which is a foundation university's preparatory school in the Mediterranean region of Turkey. Overall, 40 students and seven instructors took part in this study. These 40 students comprise 10 beginner, 16 elementary, and 14 pre-intermediate level students. As a result, the data sampling evolved into volunteer sampling in which the researcher sent a

questionnaire or a survey to target participants, and the ones who responded were considered volunteers. (Gall, Gall, & Borg, 2003). The seven instructors consist of two listening & speaking instructors, and five reading & writing instructors. The listening & speaking instructors are native speakers of English. One of them is from the UK, and the other one is from the USA. In terms of reading & writing instructors, three of them are foreigners, and two of them are Turkish. The researcher needed instructor participants' opinions and scoring as well while analyzing the data for rater reliability.

3.5. Data Collection Methods

To be able to see the impact of the end-of-year proficiency exam on students' speaking and writing skills and the impact of being tested for these skills on their opinions and emotions towards learning English, the researcher collected the video and audio files of the students' speaking tasks, the documents of students' writing tasks, and used an open ended survey. To triangulate the data and the inter-rater reliability, the instructors' comments and assessment notes on these tasks were also collected. These collected notes and comments were required to validate the interpretations of the data.

3.5.1. Collection of the Data from Students' Writing and Speaking Tasks

For the Academic Year of 2020-2021, there are 447 students enrolled in the preparatory school of the university in which the study took place. Out of 447 students, 279 students are at the beginner level, and 106 students are at the elementary level. The rest 62 students are at the pre-intermediate level. In the fall semester, the students had six tasks each for the writing and the speaking courses. Because of the online education, the schedule was covered more slowly than expected. That is why the students were only able to do five of these tasks each. The detailed information for the tasks in the fall term is given below (Table 8 & 9).

Table 8

Tasks for the Speaking Course in the Fall Term

Level	Beginner (A1)	Elementary (A2)	Pre-Intermediate (B1)
	Task 1	Task 1	Task 1
Dates & Topics	02-06 November Talk about a famous person (Individual)	02-06 November Give a presentation about an interesting place (Individual)	02-06 November Presentation- human threats to an endangered species (Individual)
	Task 2	Task 2	Task 2
Dates & Topics	16-20 November Describe photos of a place you want to visit (Individual)	16-20 November Choose an event from a festival (Group)	16-20 November Debate- New wind farm (Group)
	Task 3	Task 3	Task 3
Dates & Topics	30-04 November /December Interview students for a survey (Pair)	30-04 November /December Present a report on a device or technology (Individual)	30-04 November /December Presentation- Transport problem & solution (Individual)
	Task 4	Task 4	Task 4
Dates & Topics	14-18 December Ask for and give directions. (Pair)	14-18 December Give a presentation about climate change (Individual)	14-18 December Discussion- Are special occasions too commercial? (Group)
	Task 5	Task 5	Task 5
Dates & Topics	28- 01 December/ January Choose a person for a job. (Group)	28- 01 December/ January Have a discussion about money in sport (Group)	28- 01 December/ January Presentation- Health product or programme (Individual)

Table 9

Tasks for the Writing Course in the Fall Term

Levels	Beginner (A1)	Elementary (A2)	Pre-Intermediate (B1)
	Task 1	Task 1	Task 1
Dates & Topics	02-06 November A profile of your family	02-06 November A paragraph describing the place where you live (Its positives and negatives)	02-06 November A comparison & contrast essay (Topic free)
	Task 2	Task 2	Task 2
Dates & Topics	16-20 November A paragraph about the weather in your city or town	16-20 November A paragraph describing a festival or special event	16-20 November A cause & effect essay (Topic free)
	Task 3	Task 3	Task 3
Dates & Topics	30-04 November /December A paragraph about the life of a student in your life	30-04 November /December The internet wastes our time. It does not help us do more work. Do you agree or disagree?	30-04 November /December A problem-solution essay (Topic free)
	Task 4	Task 4	Task 4
Dates & Topics	14-18 December Facts about your country	14-18 December Write a paragraph comparing two people/places/things	14-18 December A summary paragraph & a response paragraph
	Task 5	Task 5	Task 5
Dates & Topics	28- 01 December/ January An email about a job for a friend	28- 01 December/ January Write a process paragraph (Topic free)	28- 01 December/ January An opinion essay (Topic free)

Similarly, the students are also responsible for five tasks each for both the speaking and writing courses in the spring semester. The detailed information about the tasks of the spring semester is provided in Tables 10 & 11.

Table 10

Tasks for the Speaking Course in the Spring Term

Levels	Beginner (A1)	Elementary (A2)	Pre-Intermediate (B1)
Dates & Topics	Task 1	Task 1	Task 1
	22-26 February Talk about an interesting place (Group)	22-26 February Presentation- human threats to an endangered species (Group)	22-26 February Presentation- using data from a pie chart (Individual)
Dates & Topics	Task 2	Task 2	Task 2
	08-12 March Persuade the class to go to 2 events at the National Day parades. (Group)	08-12 March Debate- New wind farm (Group)	08-12 March Decide which candidate should receive a scholarship (Group)
Dates & Topics	Task 3	Task 3	Task 3
	29-02 March/April Present a report on a device or technology (Individual)	29-02 March/April Presentation- Transport problem & solution (Individual)	29-02 March/April Role play a debate (Group)
Dates & Topics	Task 4	Task 4	Task 4
	19-23 April Give a presentation about climate change (Individual)	19-23 April Discussion- Are special occasions too commercial? (Group)	19-23 April Problem and solution presentation (Individual)
Dates & Topics	Task 5	Task 5	Task 5
	03-07 May Have a discussion about money in sport (Group)	03-07 May Presentation- Health product or programme (Individual)	03-07 May Discuss a housing problem and possible solutions (Group)

Table 11

Tasks for the Writing Course in the Spring Term

Levels	Beginner (A1)	Elementary (A2)	Pre-Intermediate (B1)
	Task 1	Task 1	Task 1
Dates & Topics	15-19 February	15-19 February	15-19 February
	Descriptive essay (Topic free)	A comparison & contrast essay (Topic free)	Explanatory essay (Topic free)
	Task 2	Task 2	Task 2
Dates & Topics	01-05 March	01-05 March	01-05 March
	Descriptive essay (Topic free)	A cause & effect essay (topic free)	A comparison & contrast essay (Topic free)
	Task 3	Task 3	Task 3
Dates & Topics	15-19 March	15-19 March	15-19 March
	Opinion essay (Topic free)	A problem-solution essay (Topic free)	Opinion essay (Topic free)
	Task 4	Task 4	Task 4
Dates & Topics	29-02 March/April	29-02 March/April	29-02 March/April
	Process essay (Topic free)	Opinion essay (Topic free)	Persuasive essay (Topic free)
	Task 5	Task 5	Task 5
Dates & Topics	12-16 April	12-16 April	12-16 April
	Narrative essay (Topic free)	Explanatory essay (Topic free)	Advantage and disadvantage essay (Topic free)

While collecting the data, the researcher tried to have a similar number of data for all levels to have a broader perspective. Hence, she needed instructors who taught at all levels. There were four speaking instructors and nine writing instructors who fit this profile. The researcher contacted them, and two of the speaking instructors and all nine writing instructors agreed to take part in this study. The researcher needed both speaking and writing tasks of the students, so the researcher needed to collaborate with only five writing instructors who shared the same classes with these two speaking instructors. Since the data were collected during the COVID-19 pandemic, all lessons were covered online. Because of this, the data were also collected online. While collecting the data in a qualitative study, ethical issues should be considered more, and some precautions should be taken in this regard (Cohen, Manion, & Marrison, 2000; Cresswell, 2013; Gall, Gall, & Borg, 2003;

Yin, 2008). To ensure the safety of the participants and not to violate their privacy, the researcher informed the participants and got their consent via consent forms. These consent forms were given to the participants online. Besides these, the researcher got the approval of the ethical committee from Gazi University and the approval of the person in charge at the university where the data were collected (In Appendix B and C, you can see the document about the approval of the ethical committee from Gazi University and the approval from the research's context). Along with these procedures the lecturers preserved the data on their computers. After all the permissions were taken, the researcher was able to access these data. The researcher had assumed that not all the students whose tasks were collected would respond to the open-ended survey, which was the second data collection tool and necessary for the triangulation of the data. Therefore, she collected the tasks from all the students whom these collaborated instructors taught. In this regard, assignments of 88 students from six different classes were collected. Among these 88 students, only 50 of them responded to the open-ended survey. Thus, the non-respondent 38 students were eliminated from the study. The other 10 participants of 50 did not answer a great number of questions in the survey. Hence, these 10 were also discarded. Therefore, the final data included 40 students' speaking and writing tasks.

The data from the collected speaking and writing tasks of participants were acquired by using rubrics, which were the same as the one that lecturers used while analyzing the tasks. These rubrics favour the analytic scoring, which provides information about different aspects of the spoken or written tasks. As an example, there are many aspects of a writing task such as its content or mechanics. A student could be effective at some parts, but s/he may not be good at other parts (Weigle, 2002). For instance, a student may be successful in writing the content, while s/he might not be effective at organization. Therefore, each aspect had better be evaluated separately to have a broader information about the students' performances. This is also applicable for the speaking task. Since the features of the mentioned rubrics corresponded to the researcher's needs, these rubrics were used in this study to see the improvement of the participants. This aforementioned rubrics were released by a well-known and prestigious international publishing house, and they have been in use for a long time. Hence, the validity and reliability of these rubrics were assumed to have been tested before. The researcher did not need to test them again. This was the another advantage of using these rubrics.

3.5.2. Collection of Responses to Open-Ended Survey

The responses to the open-ended survey were collected through a web-based survey program, which is also mentioned in Creswell's book (2012). In the literature, there are some pros and cons mentioned in terms of using a web-based survey program (Cohen, Manion, & Marrison, 2000; Creswell, 2012; Gall, Gall, & Borg, 2003). Most of these disadvantages are related to sampling, issues about the technological literacy, and the possibility of having a low number of respondents. In this study, these disadvantages were considered carefully. In terms of the first disadvantage, sampling, the target population was decided and narrowed in advance. Therefore, the survey was only sent to this population. As for the second problem concerning technological literacy, since the target population mostly consisted of generation Z, and they always use the technology due to online education, this would not be a problem. Besides, the interface of the used program was appropriate for tablets, smartphones, laptops, or desktop computers. Thanks to this, it was easy to be accessed. Last but not least, the possibility of the low number of respondents was another disadvantage. Since the researcher foresaw this problem beforehand, she collected as many tasks as possible to have the required number of respondents. Therefore, eliminating the data of the non-respondents would not be problematic. Another solution to this possibility was to write an informative cover letter. Thanks to this, there is a higher possibility to have more respondents (Creswell, 2012). On the other hand, the biggest advantage of a web-based survey is the practicality. Because of the conditions as a result of the pandemic, using a web-based survey was the only option available for the researcher.

There were two pilot studies applied before collecting the final survey data. The pre-piloting survey included questions in the target language, and there were three questions in total. This version was planned together with a colleague of the researcher and given to nine participants, three for each language level. This colleague can be considered as an expert because he holds a doctoral degree in English Language Teaching and has carried out similar studies before. In the pre-piloting process, it was discovered that the fact that questions were in the target language was the biggest problem. The respondents had difficulty in comprehending the questions and expressing themselves clearly while responding to the questions. In the first piloting period of the survey, the questions were translated into Turkish with the help of the aforementioned colleague, and there were again three questions. Once again, this version was given to the same nine respondents. In this version, the problem was about the ambiguity of the answers. Questions related to the

impact of the end-of-year proficiency exam on students' writing and speaking skills were combined into two questions. Hence, it was not clear which answer was given to which skill. The second piloting version included five questions separated for each skill and was applied to the same nine students. Eventually, there were no problems present this time, and the second piloted version was given to the respondents as the final version.

The final version of the open-ended survey consisted of four parts. The first part was the cover letter which informed the participants about the purpose of the survey and why the respondents' personal information was necessary. The researcher asked the respondents' personal information to be able to match the collected speaking and writing tasks with the responses to the survey. This first part also included a consent form. In the second part, there were two statements and one question. In the first statement, respondents' names and classes were asked. In the second statement, the e-mail addresses were asked in case the researcher needed further questions. The question in the second part was about the general impact of the end-of-year proficiency exam on students. The remaining two parts had two questions in each and four in total. The third part included two questions about the impact of the end-of-year proficiency exam on students' speaking skills. Likewise, the fourth part encompassed two questions about the impact of the end-of-year proficiency exam on students' writing skills. The open-ended survey ended by saying thank you. The survey was given to the respondents via the coursebook instructors. Each coursebook instructor shared the link of the survey in their class groups on a popular communication application. The instructors also encouraged the respondents to fill in the survey. Overall, the survey reached 88 students and 40 of them responded to the survey neatly and properly.

3.5.3. Review of the Instructors' Assessment Notes and Comments

In the literature, the researcher bias and the researchers' objectivity were two of the most common disadvantages mentioned in case studies (Adelman, Jenkins, & Kemmis, 1976; Hitchcock & Hughes, 1995; Merriam, 1998). Hence, instructors' assessment notes and comments were considered necessary by the researcher for the sake of dealing with the researcher's objectivity and the researcher's bias. Moreover, these notes and comments were required to triangulate the data. These documents were collected through e-mails and shared cloud links. They had neither been examined nor been analyzed not until the researcher had her own interpretation in order not to get affected by instructors' opinions. They were only consulted right before the researcher had the final decision about the

participants' improvement based on the analyses of the tasks, so the main purpose of the documents and the notes was to serve as a cross-check tool.

3.6. Methods for Data Analysis

The data to see the impact of end-of-year proficiency exam on students' speaking and writing tasks were collected via different tools; however, the collected tasks and survey responses were both analyzed by using the content analysis method. The review of the instructors' assessment notes and comments were examined and used as a complementary data for the triangulation of the results. While presenting the results, such abbreviations as the following were used:

Q1, Q2, Q3...: Numbers assigned to the questions in the open-ended survey

S1, S2, S3... : Numbers assigned to the students who took part in this study

CHAPTER 4

DATA ANALYSIS & FINDINGS

4.1. Introduction

In this chapter, collected data through rubrics and open-ended surveys are analyzed. Based on the analyses, the general results are presented in the tables. Detailed information about the results is provided in the further chapter.

4.2. The Analysis of the Students' Tasks and Instructors Assessment Notes and Comments

For the fall semester, three spoken and three written tasks of each student were analyzed. These tasks were Tasks 1, 3, and 5. For the spring semester, analyzing two spoken tasks (Tasks 1 & 3) for each student was enough to have clear understandings of the students' progress on speaking skills. For the analyses of written tasks during the spring semester, analyses of two written tasks (Tasks 1 & 3) for each student were enough to see 38 students' improvements in writing skills. However, for two students (S9 & S10), having a clear idea of their improvement required the analyses of one more writing task (Task 4). All the analyzed tasks were chosen randomly. In total, 200 spoken tasks and 202 written tasks were analyzed.

The analysis of the data took approximately six months. The analysis procedure started at the beginning of December, right after having the consent of the Eethical committee. It ended at the end of April. During this period, the data were analyzed three times to ensure intra-rater reliability. After the researcher had the final decision about each student's improvement, the opinions and comments of the instructors were examined to ensure inter-rater reliability. Based on these, the final results were accumulated, and they were presented in Tables 12 and 13 below.

Table 12

The Impact of End-of-Year Proficiency Exam on Students' Speaking Skills

		Content	Fluency & Coherence	Pronunciation	Grammar range & accuracy	Lexical Resource
Positive %	Improved	27,5	12,5	10	30	35
	More effective	15	35	35	37,5	30
		Content	Fluency & Coherence	Pronunciation	Grammar range & accuracy	Lexical Resource
	A little	10	17,5	17,5	25	30
	Total of Positive	%52,5	%65	%62,5	%92,5	%95
Neither positive nor negative %	Always effective	45	2,5	27,5	2,5	5
	Not improved	2,5	32,5	10	5	-
	Total of Neither positive nor negative	%47,5	%35	%37,5	%7,5	%5

Table 13

The Impact of End-of-Year Proficiency Exam on Students' Writing Skills

		Content	Mechanics (Paragrapgh)	Mechanics (Punctuation)	Organization	Grammar range & accuracy	Lexical Resource
Positive %	Improved	20	55	25	40	52,5	65
	More effective	17,5	-	-	7,5	10	7,5
	A little	20	-	2,5	15	27,5	20
Total of Positive		%57,5	%55	%27,5	%62,5	%90	%92,5
		Content	Mechanics (Paragrapgh)	Mechanics (Punctuation)	Organization	Grammar range & accuracy	Lexical Resource
Neither positive nor negative %	Always effective	35	30	57,5	25	-	-
	Not improved	7,5	15	15	12,5	10	7,5
	Total of Neither positive nor negative	%42,5	%45,5	%72,5	%37,5	%10	%7,5

Since the improvement of each student was not in the same amount, different terms were used to explain these differences. In this regard, 'Improved' is used for students who were not effective enough at the begining and got far more effective in time. 'More effective' is used to refer to the students who were effective since the beginning; however, they got more effective in time. 'A little' is used to refer to the students who were not effective enough at the beginning, and they were only able to get effective to some extent.

4.3. The Analysis of Open-Ended Surveys

The collected data through open-ended surveys were analyzed via content analysis. Since the number of participants was not high, a computer program for the analyses of qualitative data was not required. The researcher used a logbook and excel tables to keep her notes and data. Eliciting the final codes and themes took three months because the data were

analyzed three separate times in order to ensure intra-rater reliability. The first analysis took place in February. The other two occurred in March and April consecutively. Once the themes and codes consisted, and no new codes and themes emerged, the final results were ready to be presented. The set of survey questions 1, 3, and 5 aimed to find out the general impact of the end-of-year proficiency exam on students, the impact of being tested for speaking and writing skills on students' opinions and emotions towards learning English respectively. In this regard, the results related to the general impact of the end-of-year proficiency exam were presented in Table 14 below.

Table 14

The Impact of the End-of-Year Proficiency Exam on Students

Q1. Has being going to be tested in the end-of-year proficiency exam changed your emotions and opinions towards learning English? If your answer is yes, in what ways have they changed? If your answer is no, what were your original opinions and emotions?		
Codes	Themes	Percentage %
Self-awareness	Positive	50
Self- study discipline and self-improvement		
Fair decision-making system		
Just positive	Both positive and negative	5
Self-improvement		
Stress	Neutral	12,5
-		
-	Mostly Negative	7,5
Stress		
Fear of being unsuccessful	Negative	15
One exam for the decision making		
-	No idea	10

The results of the impact of being tested for speaking and writing skills on students' opinions and emotions towards learning English were presented in Tables 15 and 16 below.

Table 15

The Impact of the End-of-Year Proficiency Exam's Speaking Part on Students' Emotions and Opinions towards Learning English

Q3. The end-of-year proficiency exam tests your speaking skills. Has this situation changed your opinions or emotions towards learning English? If your answer is yes, in what ways have they changed? If your answer is no, what were your original opinions and emotions?		
Codes	Themes	Percentage %
More eager to learn	Positive	%12,5
Learning for learning – not for the exam	Neutral	%45
Being used to exams		
Fear of failure and anxiety	Negative	%42,5

Table 16

The Impact of the End-of-Year Proficiency Exam's Writing Part on Students' Emotions and Opinions towards Learning English

Q5. The end-of-year proficiency exam tests your writing skills. Has this situation changed your opinions or emotions towards learning English? If your answer is yes, in what ways have they changed? If your answer is no, what were your original opinions and emotions?		
Codes	Themes	Percentage %
More eager to learn	Positive	%30
Learning for learning – not for the exam	Neutral	%45
Fear of failure and anxiety	Negative	%25
Sad not to be able to learn		

The other two questions in the survey, Questions 2 and 4, were asked to reveal the impact of the end-of-year proficiency exam on the improvement of students' speaking and writing skills from their own perspectives respectively. The results were presented in Tables 17 and 18 below.

Table 17

The Impact of the End-of-Year Proficiency Exam on the Improvement of Students' Speaking Skills from Students' Own Perspectives

Q2. Do you believe that getting ready for the end-of-year proficiency exam's speaking part improves your speaking skills? Please explain your answers briefly.		
Codes	Answers	Percentage %
Self-improvement	Yes	%60
Self-effort		
Just enough to pass the exam		
Online education	Partially	%32,5
Thanks to self-effort only	No	%7,5
Exam-oriented		

Table 18

The Impact of the End-of-Year Proficiency Exam on the Improvement of Students' Writing Skills from Students' Own Perspectives

Q4. Do you believe that getting ready for the end-of-year proficiency exam's writing part improves your writing skills? Please explain your answers briefly.		
Codes	Answers	Percentage %
Learning for learning – not for the exam	Yes	%70
Regularly study		
Just enough to pass the exam		
Thanks to feedback		
Thanks to self-effort only	Partially	%17,5
Online education		
Only for grammar and vocabulary		
Online	No	%12,5
Exam-oriented		
More anxious		

4.4. Findings

The current study was carried out with the aim of exploring the impact of the end-of-year proficiency exam on students' speaking and writing skills along with the impact of being tested for these skills on students' opinions and emotions towards learning English at a foundation university's preparatory school in the Eastern-Mediterranean part of Turkey. To accomplish this goal, answers to seven research questions were pursued. The first two of the research questions were about the exam's structure, reliability, validity, and communicativeness. The third question was about the general impact of this proficiency exam on students. The fourth and fifth questions were related to the impact of being tested for productive skills (writing and speaking) on students' opinions and emotions towards learning English. The last two questions were about the impact of this exam on students' speaking and writing skills. The data to answer these questions were collected through students' spoken and written tasks, and an open-ended survey. The detailed information in this regard was given in the earlier sections. In the following, the findings of the study and the summary of findings are presented.

Research Question 1: Is the end-of-year English proficiency exam similar to high stake international communicative tests in terms of its structure, reliability, and validity?

To be able to answer this research question, the end-of-year proficiency exam was compared to two popular high stake international communicative tests. These were TOEFL

and IELTS. Like these exams, the end-of-year proficiency exam in the present study tests four skills. Unlike TOEFL and IELTS, the end-of-year proficiency exam tests vocabulary and grammar separately. Hence, the end-of-year proficiency exam consists of five parts. Before the pandemic of COVID-19, the exam had a different number of questions. Since this study was carried out during the pandemic, the new form was analyzed. In the reading part, there are two different texts which include about 300 words each and nine multiple-choice questions that evaluate understanding of factual information, ability to infer, vocabulary comprehension, and the connections between ideas. These texts appeal to general interest. Therefore, they do not require any prerequisite knowledge to be understood. The question types in the reading part of the end-of-year proficiency exam is similar to those in TOEFL. The time given for the reading part is 30 minutes.

In the listening part, there are two listening audios which are almost 3.30 minutes long each. These audios consist of one lecture and one conversation. The students are supposed to listen to these ideas and answer 20 multiple-choice questions which assess noticing of the important details, comprehension of the relationship between ideas, and organization of information. The given time for this part is 25 minutes. Again, the listening part is also similar to the one in TOEFL despite having different numbers of items and audios.

The speaking part is divided into three speaking tasks. In the first task, the students are expected to answer some personal questions about themselves or about their academic lifestyles. The second part includes talking about a picture. The third task is a role-play between the examiner and examinee. This role-play might be about persuading the examiner to do something or about having a discussion. The speaking part takes 10 minutes. The speaking part is similar to the speaking part of IELTS.

The writing part includes four options. The students are supposed to choose one of them and write on this topic. The given options lead the students to writing one of the essay types such as a narrative or explanatory essay. The writing part may be likened to TOEFL's first task in the writing part. The given time for the writing part is 60 minutes.

The grammar and vocabulary part includes 50 multiple-choice questions. These questions include sentence completion, sentence combination by the appropriate grammatical structure, sentence transformation, and dialogues to test vocabulary. These questions also assess students' ability to paraphrase a given sentence by using the correct grammar structure. The given time for this part is 60 minutes. This part is neither similar to IELTS nor to TOEFL, since they assess grammar and vocabulary indirectly.

TOEFL provides some statistical results for all parts, and IELTS provides statistical results for all parts except writing and speaking to show how reliable these exams are. IELTS does not provide any solid results for speaking and writing parts, for the assessment of these parts are subjective. To present the validity, both international exams show the results of studies carried out in this regard. Since the end-of-year proficiency exam's content is new due to pandemic, there are no findings in terms of validity and reliability. Thus, the researcher is not able to talk about similarities or differences, but some comments on reliability and validity are made. These comments are based on expert opinion, and on the researcher's knowledge upon, a review of related literature and similar studies.

In terms of the validity and reliability of the end-of-year proficiency exam, the reading part and the listening part include only multiple-choice questions, which increase the reliability of the test; however, this puts validity at stake to some extent. The used texts are taken from articles in real-life, so they are authentic. This authenticity contributes the degree of validity of the reading part. In real life, students may not read a text and answer multiple-choice questions. Nonetheless, they would need to read something, make inference from this text, or comprehend the factual information mentioned in the text. The audios in the listening parts consist of situations they may experience in real life. Hence, this also increases the validity of the listening part. The assessment of speaking and writing parts include a rubric of which the students have been all aware during the whole year. A rubric is a legitimate assessment tool and especially beneficial for assessing subjective and complex criteria (Muñoz & Álvarez, 2003). Using a rubric gives the opportunity of high consistency in the evaluation of students' performances. Moreover, students will know how they are going to be assessed and what standards they should meet based on the supplied expectations in the rubrics (Muñoz & Álvarez, 2003). Having an assessment criteria increases the degree of validity as well as reliability. The speaking part is assessed by two different raters. This increases the reliability of the speaking part as well for the inter-rater reliability. However, this is not applicable for the writing part. The grammar and vocabulary part consists of multiple-choice questions. Again, this is good for reliability of scoring. It is a known fact that popular international high-stakes tests do not fancy testing of grammar and vocabulary directly mainly due to the concerns related to the content validity (Hughes, 2002). These exams' contents are very wide, and sampling grammar and vocabulary in this regard is not possible. However, for the end-of- year proficiency exam studied in this study, the situation is a bit different. The large content of grammar and

vocabulary are narrowed by stating clear objectives. Furthermore, students have always been tested for grammar and vocabulary throughout the academic year. Thus, they are familiar with the question types and the content. This familiarity and students awareness on how grammar and vocabulary is tested may increase the validity to some extent.

Research Question 2: Is the end-of-year English proficiency exam communicative enough?

Like the international high-stakes proficiency tests, the exam in our study also tests four skills plus grammar and vocabulary. The audios and texts used in these exams are authentic. They include the situations students may come across in real life. Hence, it is possible to assert that reading and listening parts are communicative to some extent. The speaking part consists of tasks which let the examiners see how the examinee would act in real life if s/he faced this situation. I think this makes the speaking part communicative as well. However, one may have some doubts about the communicativeness of the writing part, and grammar and vocabulary part. In real life, the chance of writing their opinions on an issue is not very high for students, excluding their academic lives. To be more communicative, it might be better to make students address someone or read something and write a response to it in his/her writing. For the grammar and vocabulary part, it could be better to take an authentic text from an article and remove some parts to test grammar and vocabulary. In this way, the degree of communicativeness may go up.

Research question 3: How does the end-of-year proficiency exam affect the learners? Does this exam have any positive, negative, or both positive and negative impacts on learners?

Judging from the answers given to the open-ended survey, it was found that the end-of-year proficiency exam had various impacts on students. These impacts were (a) positive, (b) both positive and negative, (c) neutral, (d) mostly negative, and (e) negative. It was also found that some students had no idea about the impact of this exam on themselves. Half of the students stated that this exam had positive impact on them. Hence, the most dominant impact of this exam on students was the positive impact. The reasons why students thought so were to become more self-aware of their own progress (one out of four students), self-study and self-improvement (almost one out of five students), and to think this exam as a fair decision-making system (one student). More than half of these 20 students did not support their ideas by giving reasons. A few direct responses of the students are presented in Figure 5 below.

S4: ... because it makes me study in a more serious way.
S14: ... since it helps me to improve my English.
S23: Thanks to this exam, I will be more aware of my own skills, and I can compensate the parts I am not good at.
S28: ... because we will have to do it (knowing English) for the rest of our lives, so it is better do it during the school years.

Figure 5. Students' reasons why the exam has positive impact on them

The second dominant impact of this exam on students was the negative impact, with slightly more than one-tenth of the students. To support their ideas, they stated reasons of being stressed (half of the students) and afraid of failing (two students), and the idea of seeing this exam as an only decision-making tool (one student). The reasons why the students thought this exam affected them negatively are presented in Figure 6 below.

S11: ... The fear of failing makes me feel anxious, and being anxious might lead me to make mistakes during the exam.
S13: We have been tested several times during the year, and the final exam has the greatest percentage on passing or failing. This is not fair.
S20: To be honest, I study regularly, but I am still afraid of failing.
S26: ... Progress tests' contents are limited. However, we are responsible for all the things we have learned for a year in the final exam. This makes me feel stressed.

Figure 6. Students' reasons why the exam has negative impact on them

The other impacts of this exam on students were neutral (one out of eight students), mostly negative (a little less than one-tenth of the students), and both positive and negative (one out of twenty students). One out of ten students stated that they did not have any ideas about this exam's impacts. The students who expressed that this exam had neutral impact on them, or they did not have an idea about this exam's impact did not state any reasons. The reasons of the students who thought this exam had both positive and negative impacts on them are presented in Figure 7 below.

S10: It is positive because we will be able to know our levels. It is negative because it makes us feel stressed.

S29: ... I feel that I have to study more regularly and carefully due to this exam. This is a positive impact. The negative one is that I feel more stressful.

Figure 7. Students' reasons why the exam has both positive and negative impacts on them

Research Question 4: What do students think about the impact of being tested for the speaking skill in the end-of-year English proficiency exam on their opinions and emotions towards learning English?

To answer this research question, the responses to the related question on the open-ended survey were analyzed. Slightly less than half of the students favoured the idea that this situation had neither positive nor negative (neutral) impact on them. Slightly more than half of these students reported that they were learning for learning, and they were not learning for the exam. One out of nine students declared that they were used to exams. Five students did not state any reasons to highlight their opinions. Some reasons stated by the students who judged that this exam had neutral impact on them are shown in the Figure 8 below.

S5: I would like to learn English anyway.

S25: ... I only try to learn English to make myself a more qualified person.

S35: ... I have always favoured the idea that learning English is fun and necessary.

Figure 8. Students' reasons why being tested for speaking skills has neither positive nor negative impacts on them

With a little more than two-fifths of the students, the second-highest number was the negative impact. Among these students, nearly all of them stated the reason 'being afraid of failing and feeling anxious'. The remaining students did not mention any reasons. A few direct responses of the students why they thought so are presented in Figure 9 below.

S₁: ... It is a part of a final decision making exam. This makes me feel more stressed. It also makes learning English a more scary process.

S₂₃: Not being able to speak English makes me feel less self-confident. This discourages me to learn English.

S₃₂: I have to admit that I had had more self-confidence while speaking English before I started studying at university. Now, I feel that I do not like English as much as I used to.

Figure 9. Students' reasons why being tested for speaking skills has negative impacts on them

One out of eight students had the opinion that being tested for speaking skills had positive impacts on them towards learning English. All of them stated being more eager to learn as the reason Two of the given answers are presented in Figure 10 below.

S₁₄: I have become more ambitious.

S₃₀: I have focused on improving my speaking skills more.

Figure 10. Students' reasons why being tested for speaking skills has positive impacts on them

Research Question 5: What do students think about the impact of being tested for the writing skill in the end-of-year English proficiency exam on their opinions and emotions towards learning English?

To be able to elicit answers to this question, the responses to the related question in the open ended survey were evaluated. Upon these, it was found that most of the students (a little less than half of the students) considered that this situation had neither positive nor negative (neutral) impact on them. Almost half of the students affirmed that they were learning for learning, and they were not learning for the exam. These findings are the same as the ones found for being tested for speaking skills. A few responses are provided below (Figure 11).

S11: Writing is already a necessary skill if we want to learn English.

S17: It is exactly the same as I have expected.

Figure 11. Students' reasons why being tested for writing skills has neutral impacts on them

Three-tenths of the students expressed that being tested for writing skills had positive impacts on them. Nearly all of the students saw this as an opportunity to improve themselves, and one student became more eager to learn. Figure 12 shows some responses of the students below.

S8: I have become faster while I am writing an essay.

S21: With the help of self-effort and self-study, it definitely has a positive impact.

S33: These exams are applied to make me see my mistake and learn better. ...

Figure 12. Students' reasons why being tested for writing skills has positive impacts on them

One-fourth of students thought that being tested for writing had negative impact on them. Most of the students had the fear of failure and felt anxious. One student was sad about the possibility not to be able to learn. Although the numbers of the students were different, the stated results were the same as the ones mentioned for being tested for speaking. Some responses of the students who thought being tested for writing had negative impact on them are displayed below (Figure 13).

S3: I feel extremely anxious, and I am afraid of failing.

S23: If I could learn better, I would be happy to practice writing.

Figure 13. Students' reasons why being tested for writing skills has negative impacts on them

Research question 6: How does the end-of-year English proficiency exam affect their productive skill of speaking?

To highlight the answer to this research question, the students' own opinions about their speaking skill improvement were asked via a question in the open-ended survey. This question focused on their impressions about their own speaking skills' general improvement. The students might not fully be aware of their own progress, and they may not judge their own improvement in a detailed way. Thus, their spoken tasks were also gathered to have a clearer and safer understanding. The analyses of these tasks focused on the students' improvement in separate units. These were content, fluency and coherence, lexical resource, grammar range and accuracy, and pronunciation.

Based on the analyses of the answers to the open-ended survey, three-fifths of the students supported the idea that they improved their speaking skills thanks to getting ready for the end-of-year proficiency exam's speaking part. Almost half of them stated that the improvement was not just to pass the exam. It was general language development. Nearly one-fourth of them mentioned that they improved themselves mostly thanks to their own self-effort. One of them thought that the improvement was only enough to pass the exam. The rest of the students did not specify any reasons. When the spoken tasks of these 24 students were analyzed, mostly parallel results were found. For three-fourths of the students, the improvement was present in almost all units. However, of these 24 students, there was no improvement observed for one out of four students in the fluency unit (see Table 19 below).

Table 19

Students Who Think Getting Ready for the End-of-Year Proficiency Exam Improved Their Speaking Skills

Students	Students' ideas on their speaking skills improvement	Content	Fluency & Coherence	Lexical Resource	Grammar range & accuracy	Pronunciation
S4	yes	neutral	not improved	improved	improved	much better
S5	yes	neutral	not improved	a little	a little	neutral
S8	yes	neutral	not improved	improved	improved	much better
S10	yes	a little	not improved	a little	a little	a little
S13	yes	neutral	a little	improved	improved	a little
S14	yes	neutral	a little	improved	improved	a little
S16	yes	improved	not improved	improved	improved	much better
S18	yes	neutral	neutral	a little	a little	much better
S19	yes	neutral	much better	neutral	a little	much better
S20	yes	neutral	much better	much better	much better	much better
S21	yes	neutral	a little	a little	a little	improved
S22	yes	improved	a little	improved	a little	neutral
S27	yes	a little	much better	much better	much better	much better
S28	yes	much better	much better	much better	much better	neutral
S29	yes	much better	not improved	much better	much better	neutral
S30	yes	much better	much better	much better	much better	much better
S31	yes	much better	much better	much better	much better	much better
S32	yes	much better	much better	much better	much better	much better
S33	yes	much better	much better	much better	much better	much better
S34	yes	neutral	much better	much better	much better	much better
S36	yes	much better	much better	much better	much better	neutral
S38	yes	improved	much better	improved	improved	improved
S39	yes	neutral	not improved	a little	a little	neutral
S40	yes	improved	improved	improved	improved	neutral

Almost one-third of the students claimed that their speaking skills got better partially due to preparing themselves for the end-of-year proficiency exam's speaking part. Nearly one-fourth of them said that they could not improve themselves fully due to online education. The rest of the students did not mention any reasons. After the analyses of almost one-third of the students' spoken tasks, more than half of the elicited results were parallel with the students' own opinions. For the others, student's ideas and the results from the analyses of the spoken tasks were different. They claimed that their speaking skills partially improved. Nonetheless, development on all units of the speaking skill was observed (see Table 20 below).

Table 20

Students Who Think Getting Ready for the End-of-Year Proficiency Exam Improved Their Speaking Skills Partially

Students	Students' ideas on their speaking skills improvement	Content	Fluency & Coherence	Lexical Resource	Grammar range & accuracy	Pronunciation
S1	partially	neutral	much better	improved	improved	not improved
S2	partially	improved	improved	improved	improved	a little
S3	partially	neutral	not improved	improved	improved	a little
S6	partially	neutral	not improved	a little	a little	neutral
S7	partially	neutral	not improved	improved	improved	neutral
S9	partially	a little	not improved	a little	a little	not improved
S11	partially	improved	improved	improved	improved	a little
S12	partially	improved	improved	improved	improved	improved
S24	partially	neutral	not improved	a little	not improved	a little
S25	partially	a little not improved	a little not improved	a little	a little not improved	a little not improved
S26	partially			a little	improved	
S35	partially	improved	improved	much better	much better	neutral
S37	partially	improved	a little	improved	improved	improved

Less than one-tenth of the students asserted that getting ready for the end-of-year proficiency exam's speaking part did not improve their speaking skills. They said that they only improved themselves thanks to their own effort, and one student declared that she was not able to develop her speaking skills at all because she only concentrated on passing the exam. When the spoken tasks of these students were analyzed, it was found that these students' speaking skills improved in most of the units (see Table 21).

Table 21

Students Who Think Getting Ready for the End-of-Year Proficiency Exam Did Not Improve Their Speaking Skills

Students	Students' ideas on their speaking skills improvement	Content	Fluency & Coherence	Lexical Resource	Grammar range & accuracy	Pronunciation
S15	no	improved	much better	improved	improved	much better
S17	no	neutral	much better	improved	much better	much better
S23	no	neutral	a little	a little	a little	not improved

Research question 7: How does the final English proficiency exam affect their productive skill of writing?

The answers to this research question were elicited by asking a question in the open-ended survey and the collection of students' written tasks. The question in the open-ended survey was about students' own opinions towards their writing skills' improvement. The question in the open-ended survey focused on whether getting ready for the end-of-year proficiency exam's written part had any impact on their general written skills or not. The analyses of the tasks focused on the improvement in the task achievement, mechanics, paragraph, punctuation, lexical resource, and grammar range and accuracy. In the writing assessment rubric, mechanics unit was divided into four sub-units; punctuation, paragraphing, spelling, and capitalization. Since the students did not have any problems with spelling and capitalization, these two sub-units were not presented. To sum up, the written tasks were analyzed according to six main units and two subunits.

Upon the analyses of students' own responses, it was found that seven-tenths of these students believed that getting ready for the end-of-year proficiency exam's written part improved their general writing skills. Almost half of them said that this improvement was not just exam-oriented. It was good for their own language improvement. Nearly one fourth of the students declared that getting ready for the end-of-year proficiency exam's written part made them study regularly. Hence, they improved their writing skills. two students asserted that they improved their skills only much enough to pass the exam. It was not permanent development. One student mentioned that getting feedback during the preparation process helped her improve her writing skills. The rest of the students did not mention any specific reasons. These ideas were similar to the ones found for the impact of the preparation for the end-of-year proficiency exam's speaking part on students speaking skills. When the written tasks of the students were analyzed, parallel results were found for almost all of the students. Nevertheless, the results of the analyses of three students were different from the ones found in the analyses of the students' written tasks. Out of these three students, there were no improvements observed in any of the units for two students. For one student, the improvement was only observed in the mechanics unit. The detailed information is provided in Table 22 below.

Table 22

Students Who Think Getting Ready for the End-of-Year Proficiency Exam's Writing Part Improved Their Writing Skills

Students	Students' ideas on their writing skills improvement	Task achievement	Paragraph	Punctuation	Organization	Lexical Resource	Grammar range & accuracy
S2	yes	much better	improved	improved	improved	improved	improved
S3	yes	a little	improved	improved	improved	improved	improved
S5	yes	a little	improved	neutral	a little	a little	a little
S6	yes	not improved	improved	improved	not improved	not improved	not improved
S7	yes	much better	improved	improved	improved	improved	improved
S8	yes	much better	not improved	neutral	improved	improved	improved
S10	yes	much better	improved	neutral	improved	a little	a little
S11	yes	neutral	improved	not improved	a little	a little	a little
S12	yes	neutral	improved	neutral	improved	improved	improved
S13	yes	a little	not improved	not improved	not improved	a little	a little
S14	yes	improved	improved	improved	improved	improved	improved
S16	yes	a little	improved	not improved	not improved	a little	a little
S17	yes	improved	improved	improved	improved	improved	improved
S20	yes	improved	improved	neutral	improved	much better	much better
S21	yes	neutral	not improved	neutral	not improved	a little	a little
S22	yes	improved	improved	improved	improved	improved	a little
S24	yes	a little	improved	not improved	improved	improved	a little
S26	yes	not improved	not improved	not improved	not improved	not improved	not improved
S28	yes	neutral	neutral	neutral	neutral	improved	improved
S29	yes	neutral	neutral	neutral	neutral	improved	improved
S30	yes	neutral	neutral	neutral	neutral	improved	improved
S31	yes	neutral	neutral	neutral	neutral	improved	improved
S32	yes	much better	neutral	improved	much better	improved	improved
S33	yes	neutral	neutral	neutral	neutral	improved	much better
S34	yes	neutral	neutral	neutral	neutral	much better	much better
S35	yes	neutral	neutral	neutral	neutral	improved	improved
S38	yes	neutral	neutral	neutral	neutral	improved	improved
S39	yes	not improved	neutral	not improved	not improved	not improved	not improved

Almost one out of five students reported that preparing for the end-of-year proficiency exam's written part made their writing skills partially improved. Two students had the idea

that they improved their writing skills mostly because of their own effort. One student thought that the improvement was partly due to online education. The rest did not declare any reasons. The results from the written tasks for three students corresponded to the students' judgment on their own writing improvement. However, the results of the written tasks for four students were not parallel with the students' opinions on their own writing skills. For half of these students, there were improvements observed in all units. It was not a partial improvement. The other half of the students asserted that they were only able to develop their writing skills in terms of grammar and vocabulary. When the tasks of the other half of the students were analyzed, it was found that they improved themselves in other units as well. Contrary to what they had claimed, the improvement was not only limited to grammar and vocabulary (see Table 23 below).

Table 23

Students Who Think Getting Ready for the End-of-Year Proficiency Exam's Writing Part Improved Their Writing Skills Partially

Students	Students' ideas on their writing skills improvement	Task achievement	Paragraph	Punctuation	Organization	Lexical Resource	Grammar range & accuracy
S1	partially	improved	improved	improved	improved	improved	improved
S4	partially	neutral	improved not	improved	improved	improved	improved
S9	partially	much better	improved	neutral	improved	a little	a little
S15	partially	improved	improved	improved	improved	improved	improved
S27	partially	improved	improved	a little	neutral	improved	improved
S37	partially	much better	improved	neutral	much better	improved	improved
S40	partially	neutral	neutral	not improved	neutral	much better	much better

One out of eight students said that preparing for the end-of-year proficiency exam's written part did not improve their writing skills. Almost half of these students gave the following reasons for their position: having online education, just focusing on passing the exam, and feeling more anxious due to being tested. Each reason was stated by one different student. The rest of the students did not state any reasons. Once the written tasks of these students were analyzed, it was noticed that the opinions of the students about their own improvements and the results found according to the written tasks did not overlap. In contrast to students' own beliefs, it was discovered that the students improved in most of the units (see Table 24 below).

Table 24

Students Who Think Getting Ready for the End-of-Year Proficiency Exam's Writing Part Did Not Improve Their Writing Skills

Students	Students' ideas on their writing skills improvement	Task achievement	Paragraph	Punctuation	Organization	Lexical Resource	Grammar range & accuracy
S18	no	improved	improved	neutral	improved	a little	a little
S19	no	a little	improved	neutral	a little	improved	a little
S23	no	a little	improved	improved	a little	a little	a little
S25	no	a little	improved	neutral	a little	a little	not improved
S36	no	much better	neutral	neutral	much better	improved	improved

4.5. Summary of the Findings

In this section, findings to the research questions were presented. Firstly, the findings related to the first question “Is the end-of-year English proficiency exam’s similar to high stake international communicative tests in terms of its structure, reliability, and validity?” was presented. In order to shed light on this issue, two international communicative tests (TOEFL and IELTS) were examined, and they were compared to the end-of-year proficiency exam for their structure, reliability, and validity. In terms of the end-of-year proficiency exam’s structure, it was found that the reading, listening, and writing parts of the end-of-year proficiency exam are similar to TOEFL. The speaking part is similar to IELTS. The end-of-year proficiency exam’s grammar and vocabulary part is similar to neither TOEFL nor IELTS. Since the end-of-year proficiency exam’s format is new, no statistical results were present for reliability and validity. Therefore, the end-of-year proficiency exam could not be compared to TOEFL or IELTS. However, it would not be wrong to state that the end-of-year proficiency exam is valid and reliable to some extent based on the researcher’s comments and examination. Secondly, the findings of the second research question “Is the end-of-year English proficiency exam communicative enough?” were provided. To picture an answer to this research question, communicative tests were examined, and the literature related to communicative testing was reviewed. Based on these, it was found that the listening, reading, and speaking parts are communicative. Nevertheless, the writing part, and grammar and vocabulary part are not communicative enough. Thirdly, the findings linked with the third research question “How does the end-of-year proficiency exam affect the learners? Does this exam have any positive, negative,

or both positive and negative effects on learners?” were shown. To be able to find answers to this research question, students’ opinions were elicited through an open-ended survey. It was found that this exam had various types of impact on students; positive, both positive and negative, neutral, mostly negative, and negative. The most dominant impact on students was the positive impact, and the least dominant impact on students was both positive and negative impact. It was also found that a few students were not aware of the impact of the exam on them. Later on, the findings of the fourth research question “What do students think about the impact of being tested for the speaking skill in the end-of-year English proficiency exam on their opinions and emotions towards learning English?”, and of the fifth research question “What do students think about the impact of being tested for the writing skill in the end-of-year English proficiency exam on their opinions and emotions towards learning English?” were presented, respectively. To answer these questions, the responses were collected via an open-ended survey too. It was found that being tested for those skills had the same impact on students: positive, neither positive nor negative, and negative. On both skills, the most dominant impact was the neutral impact. The second dominant impact of being tested for writing was the positive impact whereas the second dominant impact of being tested for speaking was the negative impact. Thus, being tested for speaking affected students’ opinions and emotions towards learning English more negatively than being tested for writing did. Finally, the findings on the sixth research question “How does the end-of-year English proficiency exam affect their productive skill of speaking?”, and of the seventh research question “How does the final English proficiency exam affect their productive skill of writing?” were presented. To find answers to these research questions, students’ written and spoken tasks were analyzed besides asking students’ own opinions on their improvements in these skills. It was found that students’ opinions on their improvements and the results found upon the analyses of their tasks did not match. In contrast to students’ beliefs, the improvement in speaking skills was more than the improvement that occurred in their writing skills.

CHAPTER 5

DISCUSSION & CONCLUSION

5.1. Presentation

This study tries to highlight the impact of the end-of-year proficiency exam on students' speaking and writing skills along with the impact of being tested for those skills on students' opinions and emotions towards learning English. The discussion of the findings and conclusions drawn from them will be presented in this chapter.

5.2. Discussion

The context of the present study is limited to only one universe. Despite this fact, some general discussions and assertions can be made:

A) *The exam's stake might affect the degree of washback and impact on students' untested skills, especially on productive skills.*

The strength of the impact and washback of an exam might be affected by many factors, and the stake of the test is one of them. Based on the analyses of the collected data and the review of literature in the Turkish context, it is possible to claim that high-stake exams mostly have more negative impact on the untested skills, especially for the productive skills than the low-stake exams do. In the literature about research in Turkey, there are few studies which investigated the washback effect of high-stake exams (Çakıldere & Akpınar, 2013; Dağtan, 2012; Karabulut, 2007; Sevimli, 2007). Based on the findings of these studies except Dağtan's (2012), it would not be wrong to state that untested skills were completely ignored by the testees in the preparation period for these high-stake national tests. In Dağtan's study (2012), it was declared that test-takers included the least testing skills as well as the non-tested skills into their preparation process. In the present study, most of the student participants asserted that they were eager to improve their productive

skills in order to be more qualified people. They stated that they were not practicing these skills only for passing the exam, and Duran (2011) reported claims similar to this. The student participants in her study expressed that they would continue practicing speaking skills even if speaking was not tested.

There may be many reasons behind these findings. High-stake tests are generally result-driven, while low-stake tests are mostly process-driven. Therefore, test-takers who are supposed to take a high-stake test could be more focused on the result rather than the real language learning process. The low-stake exam studied in present research and the low-stake exam mentioned in Duran's (2011) are sort of process-driven exams. Therefore, at the micro level, learners may see this process as an opportunity to improve their general language skills to be more successful in their departments. At the macro level, being proficient in English, especially in speaking, is a requirement to be able to find a satisfactory job, and students are aware of this fact. This fact may also encourage students to take advantage of the preparatory school and improve themselves.

To sum up, these findings are worth considering since the impact and washback of low-stake exams are generally ignored in the literature both Turkey and abroad. In parallel with Shohamy's (1998) idea, it is possible to announce that low-stake exams might allow the decision-makers to yield a final high-stake decision in the long term. Hence, more emphasis should be placed on the content and administration of low-stake exams.

B) *The more valid the exam is, the more various degrees of impact the exam might have on students, and the higher the possibility of yielding positive impact is.*

Based on the analyses of the collected data, it is possible to state that the exam's content has an effect on the degrees of impact. Although the end-of-year proficiency exam has different degrees of impact on students, its general impact on students is mostly positive. Since the content areas the end-of-year proficiency exam tests are large, and the curriculum is designed to increase the development in these areas, students see this exam as an opportunity to develop their general language skills. With the testing of all the skills along with the grammar and vocabulary, students may feel more encouraged and self-confident because students' opinions and emotions towards learning English might change depending on what is tested. For example, in this study, it was discovered that being tested for writing in the proficiency exam affects students' opinions and emotions towards learning English more positively than being tested for speaking does. A student may be effective in some skills whereas s/he can be ineffective in other skills. By testing four skills

as well as grammar and vocabulary, the students' opinions and emotions towards learning English may get more positive as well. Being tested for a skill in which the student is not effective may discourage her/him from learning English. Nonetheless, this discouragement may be balanced with being tested for the skill in which s/he is effective.

Last but not least, at the macro level, testing all skills is required for communicative and valid testing while, at the micro level, it can be said that it is also important for making the students more motivated and self-confident towards learning English.

C) *The strength of impact which the exam has on the improvement of students' skills may differ in line with what is tested.*

Upon the analyses of the students' responses to the open-ended survey and students' spoken and written tasks, it was revealed that the end-of-year proficiency exam has more positive impact on students' general speaking skills than on students' writing skills. When the sub-units of these skills were analyzed, it was found that the general improvement in most students' speaking skills was more than the improvement in students' writing skills. On the other hand, the degree of impact found on two students' speaking and writing skills was very low. There could be various reasons for this, and the teachers' instructional features or the individual differences may be among these reasons. These findings were parallel with Alderson and Wall's (1993a) washback hypotheses, which stated that the washback effect of a test can be observed in some learners and some teachers. In terms of the sub-units of writing and speaking skills, the only common ones were vocabulary and grammar. It was revealed that the degree of impact on some students' improvement (three-eighths of the students) in these sub-units differed for the speaking and writing skills. In that, the degree of impact on students' grammar and vocabulary improvement was stronger in their writing skills whereas the degree of impact on students' both grammar and vocabulary or only grammar improvement was stronger in their speaking skills. Once Alderson and Wall's washback hypotheses were examined, there were not any hypotheses found to explain this situation. Hence, it is possible to come up with a new hypothesis: "The degree of impact on students' improvement may differ with respect to which skill is tested."

5.3. Conclusion

Upon examining the end-of-year proficiency exam, TOEFL, and IELTSs, similarities were found among these exams in terms of their structure. For the reliability and validity of the end-of-year proficiency exam, it may not be wrong to state that it is mostly valid and reliable. The same thing applies to its communicativeness as well. These issues are important since the structure and communicativeness of the end-of-year proficiency exam affect the impact of this exam on students.

Taking into consideration the answers given to the research questions, it is possible to declare that the prospect of being tested in the end-of-year proficiency exam and of being tested for productive skills have various impacts on students' opinions and emotions towards learning English. Firstly, the general impact of the end-of year proficiency exam is more various than the impacts of being tested for speaking and writing. The proficiency exam as a whole tests grammar, vocabulary, receptive skills as well as productive skills. Since the content is broad, and it tests different areas, the students might have separate feelings for each. When the impact of the prospect of being tested for speaking skills is compared to the impact of the prospect of being tested for writing skills, it can be seen that they both have the same impact (positive, neutral, and negative) on students. The equal number of students agreed to the idea that being tested for speaking and writing skills affected their opinions and emotions towards learning English neutrally. Besides the numbers, the reasons they gave were the same as well. They asserted that they were learning English for their own self-development. Thus, being tested for these skills did not affect their learning. These results were similar to the ones found in Duran's (2011) study in terms of speaking. She discovered that the students would continue learning speaking even if it was not tested.

In terms of the positive and negative impacts of being tested for productive skills, it was found that being tested for writing had more positive impact on students' opinions and emotions towards learning English than being tested for speaking did. The students who expressed that being tested for these skills affected their opinions and emotions towards learning English negatively enumerated similar reasons. They were afraid of failing, and they were feeling anxious. While speaking, the anxiety and fear of failure have more effect on the performance of students. On the other hand, students may get calmer and correct their mistakes more easily while writing even if they feel anxious. These might explain

why students thought being tested for speaking affected them more negatively than being tested for writing.

Secondly, the impact of the end-of-year proficiency exam on students' speaking and writing skills was analysed. It was found that the end-of-year proficiency exam affected students' writing and speaking skills either positively or neither positively nor negatively. The positive impact of the exam was observed more on students' speaking and writing skills than the neutral impact was. Students asserted that the end-of-year proficiency exam affected their writing skills more positively than it did their speaking skills. Nonetheless, it was found that the end-of-year proficiency exam had more positive impact on the students' speaking skills. It was also observed that the higher the students' proficiency levels were, the more positive impact the exam had on their speaking skills. These results were parallel with Hung and Huang's (2019) study. They stated that the link between the strength of washback and students' proficiency was evident such that the degree of washback was higher for high proficiency level students. However, this was not valid for the impact on students' writing skills. As for writing, the degree of impact was stronger when students' proficiency levels were lower. The neutral impact was more dominant on students' writing skills from higher proficiency levels. The students who were at high proficiency levels were already effective in a few units such as content and paragraphing even at the beginning. The neutral impact was more dominant in these areas. Hence, this might lead to the observation that there is more positive impact on students' writing skills from lower levels.

Last but not least, this end-of-year proficiency exam is a low-stake exam, but it has different degrees of impact on students' opinions and emotions towards learning English as well as on their speaking and writing skills. No matter how effective an education system is, the impact of the exam on students' learning and improvement had better not be ignored. Regardless of most students' statements about learning for learning but not for the exam, it would be better to follow communicative teaching and testing. Additionally, an appropriate guidance during the preparation of the exam should be emphasized to yield more positive impact and washback.

5.4. Suggestions for Further Research

The current study was carried out to see the impact of the end-of-year proficiency exam on students' speaking and writing skills as well as the impact of being tested for these skills

on students' opinions and emotions towards learning English. The study was conducted during the pandemic of COVID-19. Therefore, the online education was present, and the administration and the content of the exam were different from the face-to-face education as well. Because of the limited time, the proficiency exams' results of students could not be analyzed and compared with their improvement. Lastly, the validity and reliability of the end-of-year proficiency exam could not be calculated and presented statistically due to the researcher's lack of experience. If the study was to be carried out again, these issues can be taken into consideration to have a deeper and clear understanding of the case.

The following recommendations for further research can be made:

1. This study was conducted during the pandemic of COVID-19. This affected the content and the administration of the exam. The study should also be conducted during face-to-face education since the impact of the end-of-year proficiency exam during face-to-face education needs to be understood.
2. It would be better to see the students' results of the end-of-year proficiency exam; however, it was not possible due to the time limit. A follow-up study should be carried out to see the students' performance in the end-of-year proficiency exam.
3. The validity and the reliability of the end-of-year proficiency exam were presented based on the researcher's comments. They should be also analyzed and presented professionally to have a clearer picture of the validity and reliability of this exam.
4. This study only focused on the impact of the end-of-year proficiency exam on students' productive skills. If this study was to be done again, the impact of the end-of-year proficiency exam on receptive skills could be investigated, and the comparison between the impact of this exam on student's receptive and productive skills could be made as well.

5.5. Recommendations for Teachers

Although this study was carried out at the preparatory school of a foundation university, the following suggestions could be helpful for teachers who work at the preparatory schools of other universities.

1. A great number of students think that improving their writing and speaking skills is not just a requirement to pass the exam. They see this process as a tool to improve their

language skills in general. Therefore, teachers should be aware of this situation and try not to teach exam-oriented merely. The teachers who mainly teach language areas like grammar and vocabulary also need to be aware that they can help improve these skills.

2. Feedback plays an important role in students' improvement of speaking and writing skills. Thus, the importance of giving feedback should not be ignored, and students had better be supplied with regular and clear forms of feedback.

5.6. Recommendations for Testing Units

The following suggestions could be helpful for teachers who work at the testing units of the preparatory schools of other universities:

1. The structure and the communicativeness of the exam are crucial to yield positive impact on students' speaking and writing skills. Therefore, it is necessary to make well-structured and communicative tests.
2. Including four skills in the progress tests or giving tasks to practice these skills regularly may increase the chance of yielding positive impact.
3. Stating the assessment criteria might increase the possibility of beneficial impact and washback by making students aware of what is expected from them and what the standards are.

REFERENCES

- Adelman, C., Jenkins, D., & Kemmis, S. (1976). Re-thinking case study: notes from the second Cambridge Conference. *Cambridge Journal of Education*, 6(3), 139-150.
- Akpınar, K.D. & Çakıldere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and ÜDS) on productive and receptive skills of academic personnel. *Journal of Language and Linguistic Studies*, 9(2), 81-94
- Alderson, I. C., & Walt, D. (1993a). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Alderson, I. C., & Walt, D. (1993b). Examining washback: The Sri Lankan impact study. *Language Testing*, 10, 41–69.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Brown, J. D. (1996). *Testing in language programs*. New Jersey: Prentice Hall Regents.
- Brown, H. D. (2004). *Language assessment: principles and classroom practices*. New York: Longman.
- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Research Contexts and Methods* (pp. 113–128).

- Burrows, C. (1998). Searching for washback: An investigation into the impact on teachers of the implementation into the adult migrant English program of the certificate in spoken and written English. in Cheng, L., & Watanabe, Y. (eds), *Research Contexts and Methods* (pp. 113-128).
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carr, N. (2011). *Designing and Analyzing Language Tests*. Oxford: Oxford University.
- Cheng, L. Curtis, A. (2012). Test impact and washback: Implications for teaching and learning. In C. Coombe, P., Davidson, B. O' Sullivan and S. Stoyynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 89-95). Cambridge: Cambridge University.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing testing on teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-17). Mahwah: Lawrence Erlbaum Associates.
- Cheng, L. E. & Watanabe, Y. E. (2004). *Washback in language testing: Research contexts and methods*. Lawrence Erlbaum Associates.
- Cheng, L. (1999). Changing assessment: Washback on teacher perspectives and actions. *Teaching and Teacher Education*, 15, 253–271.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11, 38–54.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. Routledge.
- Dağtan, E. (2012). *An examination on the wasback effect of the Interuniversity Board Foreign Language Test (The example of Dicle University)*. (Unpublished master's thesis). Dicle University, Turkey.
- Dornyei, Z. (2007) *Research methods in applied linguistics*. Oxford: Oxford University.
- Duran, Ö. (2011). *Teachers' and students' perceptions about classroom-based speaking tests and their washback*. (Unpublished master's thesis). Bilkent University, Ankara.

- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London & New York: Routledge.
- Gall, M. D., Gall, J. P., & Walter, B. R. (2014). *Applying Educational Research: How to read, do, and use research to solve problems of practice*. London: Pearson.
- Gall, J. P., Gall, M. D., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston: Allyn and Bacon.
- Gates, S. (1995). Exploiting washback from standardized tests. *Language Testing in Japan*, 107-112
- Green, A. (2013). *Exploring language assessment and testing: Language in action*. London: Routledge.
- Green, A. (2006). Watching for washback: observing the influence of the international english language testing system academic writing test in the classroom. *Language Assessment Quarterly*, 3(4), 333-368.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295-303.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. in Cheng, L., & Watanabe, Y. (eds.), pp. 97-111.
- Heaton, J.B. (1990). *Writing English language tests*. London and New York: Longman.
- Hitchcock, G., & Hughes, D. (1995). *Research and the teacher*. London: Routledge.
- Hughes, A. (2002). *Testing for language teachers*. Cambridge: Cambridge University.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. (Unpublished manuscript). University of Reading.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University.
- Hung, S. A., & Huang, H.D. (2019). Standardized proficiency tests in a campuswide English curriculum: A washback study. *Language Testing in Asia*, 9(21), 1-17.
- Karabulut, A. (2007). *Micro level impacts of foreign language test (university entrance examination) in Turkey: A washback study*, (Unpublished master's thesis). Iowa State University, USA.

- Köktürk, S. (2015). *Washback effects of examinations on teaching and learning English in Gazi University Preparatory School*, (Unpublished Master's thesis). Gazi University, Ankara.
- Lexico Online Dictionary. Backwash. In Lexico.com dictionary. Retrieved from <https://www.lexico.com/definition/backwash>
- Merriam, S. B. (1988) *Case study research in education*. San Francisco, CA: Jossey Bass.
- Merriam-Webster. (n.d.). Proficient. In *Merriam-Webster.com dictionary*. Retrieved from <https://www.merriamwebster.com/dictionary/proficient>
- Messick, S. (1996). Validity and washback in language testing. *Language testing*, 13(3), 241-256.
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 1-17.
- Muñoz, A. P., & Álvarez, M. E. (2003). Estimating the validity and reliability of an oral assessment instrument. *REVISTA Universidad EAFIT*, 39(132), 65-75.
- Nitko, A., & Brookhart, S. (2013). *Educational Assessment of Students*. Pearson New International Edition.
- Özmen, K. S. (2011a). Analyzing washback effect of SEPPPO on prospective English teachers. *The Journal of Language and Linguistic Studies*, 7(2), 24-52.
- Özmen, K. S. (2011b). Washback effect of the inter-university foreign language examination on foreign language competences of candidate academics. *Novitas-ROYAL (Research on Youth and Language)*, 5(2), 215-228.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback of high-stakes test. *Language Testing*, 22(2), 142-173.
- Sadler, D. R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, 54, 60-79.
- Sarı, S. Y. (2018). *The impact of institutional tests on teachers' teaching methodologies* (Unpublished master's thesis). Çağ University, Mersin.
- Sato, T. (2019). An investigation of factors involved in Japanese students' English learning behavior during test preparation. *Papers in Language Testing and Assessment*, 8(1), 69-95.

- Sevimli, S. (2007). *Washback effects of foreign language component of the university entrance examination on the teaching and learning context of English language groups in secondary education*, (Unpublished master's thesis). University of Gaziantep, Turkey.
- Sheriff, A. (2019). *Proficiency Test Reviews: the TOEFL iBT, IELTS, and PTE-Academic*. 2019-2020 Catalog of Colorado State University.
- Shohamy, E. (2001). *The power of test: A critical perspective on the uses of language tests*. London: Longman.
- Shohamy, E. (1998). Inside the "black box" of classroom language tests. *Studies Anglica Posnaniensia* 33, 343-352. (special issue in honor of Kari Sajaavara).
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Shih, C. M., (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135-162.
- Stake, R. (2000). Qualitative case studies. In Denzin, K. N., & Lincoln, Y. S. (Eds.). *The Sage handbook of qualitative research* (pp. 443-466). London: Routledge.
- Tsagari, D. (2007). *Review of washback in language testing: How has been done? What more needs doing?* (Unpublished Paper). United Kingdom; Lancaster University.
- Wall, D. (2012). Washback in language assessment. *The Encyclopedia of Applied Linguistics*, Edited by Carol A. Chapelle (2013). New Jersey: Blackwell.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge: Cambridge University.
- Wall, D. (1997). Impact and washback in language testing. *Encyclopedia of Language and Education*, 7, 291-302.
- Watanabe, Y. (1996). Investigating washback in Japanese EFL classrooms: Problems and methodology *Australian Review of Applied Linguistics*, 13, 208-239.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University.
- Wittgenstein, L. (2010). *Tractatus logico-philosophicus*. Project Gutenberg e-book.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.



APPENDICES



Appendix A. Open-ended Survey for Students

Questions in Turkish:

1. Yıl sonunda yapılan İngilizce yeterlilik sınavında test edilecek olmanız İngilizce öğrenmeye karşı duygu ve düşüncelerinizi değiştirdi mi? Eğer cevabınız evet ise bu değişiklikler ne yönde oldu? Eğer cevabınız hayır ise orijinal duygu ve düşünceleriniz nelerdi?
2. Yıl sonunda yapılan İngilizce yeterlilik sınavının konuşma kısmına hazırlanmanın İngilizce konuşma becerilerinizi geliştirdiğini düşünüyor musunuz? Kısaca nedenleriyle açıklayınız.
3. Yıl sonunda yapılan İngilizce yeterlilik sınavında konuşma becerilerinizin test edilecek olması İngilizce öğrenmeye karşı duygu ve düşüncelerinizi değiştirdi mi? Eğer cevabınız evet ise bu değişiklikler ne yönde oldu? Eğer cevabınız hayır ise orijinal duygu ve düşünceleriniz nelerdi?
4. Yıl sonunda yapılan İngilizce yeterlilik sınavının yazma kısmına hazırlanmanın İngilizce yazma becerilerinizi geliştirdiğini düşünüyor musunuz? Kısaca nedenleriyle açıklayınız.
5. Yıl sonunda yapılan İngilizce yeterlilik sınavında yazma becerilerinizin test edilecek olması İngilizce öğrenmeye karşı duygu ve düşüncelerinizi değiştirdi mi? Eğer cevabınız evet ise bu değişiklikler ne yönde oldu? Eğer cevabınız hayır ise orijinal duygu ve düşünceleriniz nelerdi?

Questions in English:

1. Has being going to be tested in the end-of-year proficiency exam changed your emotions and opinions towards learning English? If your answer is yes, in what ways have they changed? If your answer is no, what were your original opinions and emotions?
2. Do you believe that getting ready for the end-of-year proficiency exam's speaking part improves your speaking skills? Please explain your answers briefly.
3. The end-of-year proficiency exam tests your speaking skills. Has this situation changed your opinions or emotions towards learning English? If your answer is yes, in what ways have they changed? If your answer is no, what were your original opinions and emotions?

4. Do you believe that getting ready for the end-of-year proficiency exam's writing part improves your writing skills? Please explain your answers briefly.
5. The end-of-year proficiency exam tests your writing skills. Has this situation changed your opinions or emotions towards learning English? If your answer is yes, in what ways have they changed? If your answer is no, what were your original opinions and emotions?



Appendix B. The Approval of Gazi University's Ethical Committee



T.C.
GAZİ ÜNİVERSİTESİ



Sayı : E-91610558-302.08.01-
Konu :

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi :

İlgi yazınız ile göndermiş olduğunuz, Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı, İngilizce Dili Eğitimi Bilim Dalı Yüksek Lisans Öğrencisi Rukiye Elifnur ÖZMEN'in, Doç.Dr.Cemal ÇAKIR'ın danışmanlığında yürüttüğü "*Impact of Preparatory School's Final English Proficiency Exam on Students' Speaking and Writing Skills*" adlı tez çalışması ile ilgili konu Kurulumuzun tarih ve sayılı toplantısında görüşülmüş olup,

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

Bilgilerinizi ve gereğini rica ederim.

Araştırma Kod No:



Ek:1Liste

Belge Doğrulama Kodu :BE8VA0M3T

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Takip Adresi : <https://belgedogrulama.gazi.edu.tr/belgedogrulama.aspx>



Bu belge, güvenli elektronik imza ile imzalanmıştır.

Appendix C. The Permission of the University Where the Study Took Place



Doç.Dr. Cemal Çakır'ın sorumlu araştırmacısı olduğu Rukiye Elifnur Özmen adlı kişinin araştırmacı olduğu Impact of Preparatory School's Final English Proficiency Exam on Students' Speaking and Writing Skills isimli araştırma projesi kapsamında [redacted] yapılan derslerin video ve ses kayıtlarının kullanılacağı konusunda bilgilendirildim.

Bilgilerinize arz ederim.

Tarih: [redacted]

Kurum: [redacted]

Unvan: [redacted]

Görev: [redacted]

Ad ve Soyad: [redacted]

İmza: [redacted]



GAZİLİ OLMAK AYRICALIKTIR..