



**HOW WASHBACK OPERATES IN MACRO- AND MICRO-LEVEL
DYNAMICS OF EFL TEACHING IN TURKEY: INSIGHTS FROM
EFL TEACHERS' TESTING BELIEFS, LITERACY AND
PRACTICES**

Fatıma Nur Fişne

DOCTORAL DISSERTATION

**DEPARTMENT OF FOREIGN LANGUAGES EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM**

GAZİ UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

MARCH, 2022

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 12 ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Fatıma Nur
Soyadı : Fişne
Bölümü : İngiliz Dili Eğitimi
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı : Türkiye’de Ölçme ve Değerlendirme Etkisi İngilizce Öğretiminin Makro ve Mikro Dinamiklerinde Nasıl Faaliyet Göstermektedir: İngilizce Öğretmenlerinin Ölçme ve Değerlendirme İnançları, Okuryazarlıkları ve Uygulamalarına Yönelik Görüşler

İngilizce Adı : How Washback Operates in Macro- and Micro-Level Dynamics of EFL Teaching in Turkey: Insights From EFL Teachers’ Testing Beliefs, Literacy and Practices

ETİK İLKELERE UYGUNLUK BEYANI

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

Yazar Adı Soyadı: Fatıma Nur FİŞNE

İmza:

JÜRİ ONAY SAYFASI

Fatıma Nur Fişne tarafından hazırlanan “How Washback Operates in Macro- and Micro-Level Dynamics of EFL Teaching in Turkey: Insights From EFL Teachers’ Testing Beliefs, Literacy And Practices” adlı tez çalışması aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Gazi Üniversitesi İngiliz Dili Eğitimi Anabilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Kemal Sinan ÖZMEN

İngiliz Dili Eğitimi Anabilim Dalı, Gazi Üniversitesi

Başkan: Prof. Dr. Abdulvahit ÇAKIR

İngiliz Dil Bilimi Anabilim Dalı, Ufuk Üniversitesi

Üye: Prof. Dr. Paşa Tefvık CEPHE

İngiliz Dili Eğitimi Anabilim Dalı, Gazi Üniversitesi

Üye: Prof. Dr. Cem BALÇIKANLI

İngiliz Dili Eğitimi Anabilim Dalı, Gazi Üniversitesi

Üye: Prof. Dr. Çiler HATİPOĞLU

İngiliz Dili Eğitimi Anabilim Dalı, Orta Doğu Teknik Üniversitesi

Tez Savunma Tarihi: 25/02/2022

Bu tezin İngiliz Dili Eğitimi Anabilim Dalında Doktora tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Yücel GELİŞLİ

Eğitim Bilimleri Enstitüsü Müdürü



To my beloved mother, Saliha FİŞNE...

ACKNOWLEDGEMENT

Firstly, I would like to express my profound gratitude to my supervisor, Prof. Dr. Kemal Sinan ÖZMEN, who helped my dream of researching washback come true and guided me a lot in the design, implementation and reporting stages of my PhD dissertation. Without his encouragement and support, I would have never completed this dissertation.

Next, I would like to extend my thanks to Prof. Dr. Abdulvahit ÇAKIR, whose knowledge of language testing I always admire, Prof. Dr. Paşa Tefik CEPHE, who supported my dissertation studies all the time, Prof. Dr. Cem BALÇIKANLI, who made great contributions to my research procedures, and Prof. Dr. Çiler HATİPOĞLU, whose thought-provoking feedback made my dissertation stronger.

I collected the first part of my research data from the Ministry of National Education, where I began teaching in 2012 and aimed at contributing to language testing. This motivation triggered me to interview with EFL teachers, and I really thank each language teacher for allocating time to enhance my study. In a similar way, I am very grateful to field experts who enlightened me about the research constructs and helped me make sense of the findings.

In addition, I would like to thank Res. Assist. Merve ÖKSÜZ ZEREY for her endless support that inspired me to work harder and complete my study. Furthermore, many thanks go to my colleagues who are the researchers in the fields of English Language Teaching and Educational Measurement and Evaluation at Gazi University and made valuable comments on the validity of my interview questions. I also express my thanks to my dear friends Serap ASLAN and Fadime AKSAK for their friendship that is always a source of motivation for me to study further.

What's more, I want to thank my family members, *my mother, father, three brothers, sister-in-law* and *niece*, wholeheartedly because I would have never finalized my dissertation without their support, encouragement, and patience.

Finally, I would like to acknowledge my gratitude to TÜBİTAK for supporting me to carry out my PhD research study within the program of TÜBİTAK 2211-E Scholarship for Doctoral Programs.



**HOW WASHBACK OPERATES IN MACRO- AND MICRO-LEVEL
DYNAMICS OF EFL TEACHING IN TURKEY: INSIGHTS FROM
EFL TEACHERS' TESTING BELIEFS, LITERACY AND
PRACTICES**

(Doctoral Dissertation)

Fatma Nur Fişne
GAZİ UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
March 2022

ABSTRACT

Language testing is an essential component of English language teaching and influences teaching dynamics in different ways. The effect of language tests on teaching or learning is defined as washback effect. Washback is a multidimensional and complex phenomenon which takes a pivotal role in shaping low-stakes or high-stakes decisions on instructional dynamics, stakeholders, educational system and wider community. There are also some operative forces that mediate the dimensions of the washback effects. This study aims to explore the washback effects of English language testing on teaching dynamics in Turkey and then macro- and micro-level triggering factors. To this end, it adopts a case-study research design within the framework of qualitative research methodology. There are three cases (i.e., primary, lower secondary and upper secondary education) and two units of analysis (i.e. teacher-based assessments and centralized tests) in each case. Qualitative data were gathered from 25 EFL teachers who work in MoNE and 9 field experts who have seminal works about language teaching and testing policies and practices. Semi-structured

interviews were used as the data collection tools, and interview data were analyzed through content analysis method. The study findings were examined in terms of the dynamics of each case. For the primary education, the 4th grade formal exams cause a methodological shift in language assessment. In some school settings, these exams may trigger exam oriented-teaching and test-taking anxiety; however, EFL teachers' missions and beliefs play a mediating role in the operationalization of these washback effects. On the other hand, these exams give feedback on language learning. Another important finding in this case is the use of mock tests and demand on practicing test items. Socio-economic concerns and centralized testing system are macro-level triggers that are influential in promoting these practices. As for lower secondary education, it is seen that teacher-based assessment, mock tests and high-stakes tests influence language teaching dynamics. Teacher-based assessment provides a basis for predicting further success, proving achievement, and giving feedback on learning. However, it causes exam anxiety triggered by high-stakes testing policies. Similarly, mock tests exert impacts on teaching and learning. Regarding the washback effects of USE-EE, the basic findings are methodological shifts in both teaching and assessment and development of some instructional policies. High-stakes testing policies, socio-economic dynamics and teacher education policies are the macro-level triggers that influence micro-level triggers and facilitate the abovementioned washback effects. Finally, upper secondary education is the case that holds the use of teacher-based assessment and two high-stakes tests: BCT and FLT. Regulation-triggered low-stakes skills-based assessment basically influences teacher behaviours and mastery of language skills. From high-stakes perspectives, BCT and FLT have various washback effects. Since no language item is included in BCT, it demotivates non-major English learners to practice English language. This also affects teacher motivation negatively. On the other side, FLT causes a sharp transition to teaching grammar topics and strategies covered in this language test. The main triggering factors influential in shaping these effects are holistically employability and high-stakes testing policies.

Key words: Language testing; washback effect; assessment practices

Page Number: xix + 182

Supervisor: Prof. Dr. Kemal Sinan ÖZMEN

**TÜRKİYE’DE ÖLÇME VE DEĞERLENDİRME ETKİSİ
İNGİLİZCE ÖĞRETİMİNİN MAKRO VE MİKRO
DİNAMİKLERİNDE NASIL FAALİYET GÖSTERMEKTEDİR:
İNGİLİZCE ÖĞRETMENLERİNİN ÖLÇME VE
DEĞERLENDİRME İNANÇLARI, OKURYAZARLIKLARI VE
UYGULAMALARINA YÖNELİK GÖRÜŞLER**

(Doktora Tezi)

**Fatıma Nur Fişne
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

Mart 2022

ÖZET

Ölçme ve değerlendirme, İngilizce öğretiminin önemli bir parçasıdır ve öğretim dinamiklerini farklı şekillerde etkilemektedir. Dil sınavlarının öğretime olan bu etkisine “ölçme ve değerlendirme etkisi” denilmektedir. Ölçme ve değerlendirme etkisi, öğretim dinamikleri, paydaşlar, eğitim sistemi ve toplum ile ilgili kararların şekillenmesinde önemli bir rol oynayan çok boyutlu ve karmaşık bir olgudur. Ölçme ve değerlendirme etkisinin boyutları üzerinde faaliyet gösteren bazı etkin faktörler vardır. Bu çalışma, ölçme ve değerlendirmenin İngilizce öğretimi üzerine olan etkilerini ve bu etkileri meydana getiren makro ve mikro düzey faktörleri keşfetmeyi hedeflemektedir. Bu amaç doğrultusunda, nitel

araştırma yöntemi çerçevesine durum çalışması araştırma deseni benimsenmiştir. Araştırmada ilkokullar, ortaokullar ve liseler olmak üzere üç temel durum ve öğretmen temelli ölçme ve merkezi sınavlar olmak üzere her bir durum için iki analiz birimi bulunmaktadır. Nitel veri, MEB’de görev yapan 25 İngilizce öğretmeninden ve dil öğretimi, ölçme ve değerlendirilmesi politikaları ve etkinlikleri alanında önemli çalışmalara sahip olan 9 alan uzmanından toplanmıştır. Veri toplama aracı olarak yarı-yapılandırılmış görüşme soruları kullanılmıştır. Görüşme verileri, içerik analizi yöntemi ile çözümlenmiştir. Araştırma bulguları, her bir durumun kendi dinamiklerine göre incelenmiştir. İlkokullar için, 4. Sınıflarda uygulanan yazılı sınavlar ölçme ve değerlendirmede yöntemsel bir değişikliğe sebep olmaktadır. Bazı okullarda, bu sınavlar sınav odaklı öğretime ve sınav kaygısına neden olmaktadır fakat öğretmen inançları ve amaçları bu etkilerin ortaya çıkmasında arabuluculuk görevi üstlenmektedir. Diğer taraftan, bu sınavların öğrenmeye yönelik dönüt verdiği görülmüştür. Bir diğer önemli bulgu ise deneme sınavları ve test kullanımına yönelik taleptir. Sosyo-ekonomik kaygılar ve merkezi sınav sistemi bu talepleri etkileyen makro düzey faktörlerdir. Ortaokullarda, öğretmen temelli ölçme, deneme sınavı ve önemli sonuçları olan sınavlarının dil öğretim dinamiklerini etkilediği görülmektedir. Öğretmen temelli ölçme, ileriye yönelik başarıyı tahmin etmek, başarıyı kanıtlamak ve öğrenme ile ilgili dönüt vermek için temel teşkil etmektedir. Öte yandan, önemli sonuçları olan sınav politikaları nedeni ile öğrencilerde sınav kaygısına da sebep olmaktadır. Benzer şekilde, deneme sınavları öğretim ve öğrenme üzerinde bazı etkiler meydana getirmektedir. Liselere Geçiş Sınavı (LGS) göz önüne alındığında, temel bulgular bu sınavın öğretim ve ölçmede yöntemsel değişiklikler meydana getirdiğini ve bazı politikaların gelişmesine neden olduğunu göstermektedir. Önemli sonuçları olan sınav politikaları, sosyo-ekonomik dinamikler ve öğretmen eğitimi politikaları mikro-düzen faktörleri etkilemekte ve söz konusu etkilerin meydana gelmesine neden olmaktadır. Son olarak, liseler TYT ve YDT olmak üzere iki adet önemli sonuçları olan sınavı ve öğretmen temelli ölçmeyi kapsayan bir araştırma durumudur. Yönetmelikler tarafından yönlendirilen beceri odaklı sınavlar temel olarak öğretmen davranışlarını ve dil becerilerinin kazanım derecelerini etkilemektedir. Önemli sonuçları olan sınavlar tarafından bakıldığında ise bu sınavların farklı etkilerinin olduğu görülmektedir. TYT’de İngilizce sorusunun olmaması dil sınıfında olmayan öğrencilerin İngilizce öğrenme motivasyonlarını etkilemektedir. Bu durum öğretmen motivasyonunu da olumsuz etkilemektedir. Diğer taraftan, YDT, testte yer alan dil bilgisi konularının ve stratejilerinin öğretimi noktasında büyük bir değişikliğe neden olmaktadır. Bu etkileri ortaya çıkaran ana faktörle istihdam politikaları ve önemli sonuçları olan ölçme ve değerlendirme politikalarıdır.

Anahtar Kelimeler: Dil ölçme ve değerlendirme; ölçme değerlendirme etkisi; ölçme ve değerlendirme uygulamaları

Sayfa Sayısı: xix + 182

Tez Danışmanı: Prof. Dr. Kemal Sinan ÖZMEN

TABLE OF CONTENTS

ABSTRACT.....	vii
ÖZ	ix
LIST OF TABLES	xvi
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS.....	xix
CHAPTER I	1
INTRODUCTION.....	1
Problem Statement.....	1
Aim of the Study.....	4
Research Questions	5
Significance of the Study.....	6
Assumptions of the Study.....	7
Limitations of the Study.....	7
Definitions.....	8
CHAPTER II.....	11
REVIEW OF LITERATURE.....	11
Basic Concepts in Language Testing.....	11
Test Purposes and Classifications.....	13
Qualities of Language Tests	14
Washback	16

Research Studies on Washback in International Contexts.....	19
Research Studies on Washback in Turkish Context.....	24
Critical Language Testing Theory.....	27
Language Assessment Literacy and Beliefs.....	28
Educational Policies in Turkey	29
The Turkey K-12 Education System: Language Teaching and Testing	33
<i>Primary and Lower Secondary Education</i>	33
<i>Upper Secondary Education</i>	36
CHAPTER III.....	41
METHODOLOGY.....	41
Research Aims and Questions.....	41
Research Design	42
Sampling.....	45
Participants	46
EFL Teachers.....	46
Field Experts	50
Data Collection	51
Data Collection Tools.....	52
<i>Interview Form for EFL Teachers</i>	52
<i>Interview Form for Field Experts</i>	55
Data Analysis.....	56
Reliability and Validity.....	59
CHAPTER IV.....	65
FINDINGS.....	65
Case 1: Primary Education.....	65
Washback Effects of Formal Written Examinations and Triggering Factors	66

<i>Methodological Shift in Language Assessment</i>	66
<i>Exam-Oriented Language Teaching</i>	68
<i>Feedback on Language Learning</i>	69
<i>Exam Anxiety</i>	69
Use of Mock Exams and Practice Tests	72
<i>Mock Examination Washback and Triggering Factors</i>	72
<i>Washback Effects of Practice Tests and Triggering Factors</i>	75
Teacher Beliefs and Missions	79
Case 2: Lower Secondary Education	80
Washback Effects of Low-Stakes Assessment and Triggering Factors	81
<i>Teacher-Based Assessment Washback and Triggering Factors</i>	81
<i>Mock Examination Washback and Triggering Factors</i>	84
Washback Effects of High-Stakes Testing and Triggering Factors	86
<i>USE-EE Washback on Language Teaching, Learning and Teachers</i>	86
<i>Methodological Shift in Language Teaching</i>	86
<i>Use of Supplementary Materials</i>	88
<i>Teacher Behaviours</i>	89
<i>Exam-Driven Learning Expectations</i>	90
Washback Effects of USE-EE on EFL Assessment Practices	91
<i>Methodological Shift in Language Assessment</i>	91
<i>Aligning Language Assessment with USE-EE</i>	93
Washback Effects of USE-EE on Policy-Making	94
Triggering Factors for Washback Effects of USE-EE	98
<i>Stakeholders' Perceptions and Expectations</i>	99
<i>Socio-Economic Dynamics</i>	102

<i>Institutional Dynamics</i>	<i>103</i>
<i>Low-Stakes Assessment Policies</i>	<i>104</i>
<i>Teaching Duration</i>	<i>106</i>
<i>Teacher Professional Identity and Assessment Literacy.....</i>	<i>106</i>
Case 3: Upper Secondary Education	110
Washback Effects of Low-Stakes Assessment and Triggering Factors	110
<i>Washback Effects of Teacher-Based Assessment</i>	<i>110</i>
<i>Enhancement of Spoken Production</i>	<i>110</i>
<i>Feedback on Language Learning</i>	<i>111</i>
<i>Affective Outcomes.....</i>	<i>112</i>
<i>Exam-Oriented Learning Habits</i>	<i>113</i>
<i>Teacher Collaboration</i>	<i>113</i>
<i>Teacher Autonomy</i>	<i>115</i>
Triggering Factors for Teacher-Based Assessment Washback.....	115
<i>High-Stakes and Low-Stakes Assessment Policies</i>	<i>115</i>
Washback Effects of High-Stakes Assessment and Triggering Factors	117
<i>Washback Effects of BCT on Language Learners and Teachers</i>	<i>118</i>
<i>Motivation Loss.....</i>	<i>118</i>
Washback Effects of FLT on Language Teaching and Learning	119
<i>Methodological Shift in Language Teaching.....</i>	<i>119</i>
<i>Teaching Test Strategies</i>	<i>121</i>
<i>Mastery of Language Skills</i>	<i>122</i>
Washback Effects of FLT on Policy-Making	123
<i>Teacher Designation</i>	<i>123</i>
Washback Effects of FLT on Language Assessment.....	124

<i>Methodological Shift in Language Assessment</i>	<i>124</i>
<i>Exam-Oriented Language Assessment</i>	<i>126</i>
<i>Stakeholders' Perceptions and Expectations</i>	<i>129</i>
<i>Teacher Beliefs, Practices and Assessment Literacy</i>	<i>139</i>
CHAPTER V	143
DISCUSSION	143
Case 1: Primary Education.....	143
Case 2: Lower Secondary Education.....	145
Case 3: Upper Secondary Education	147
CHAPTER VI.....	151
CONCLUSION AND IMPLICATIONS.....	151
Conclusion	151
How Washback Effects Operate in EFL Teaching Dynamics in Turkey	151
Implications	154
REFERENCES.....	155
APPENDICES	167

LIST OF TABLES

Table 1. <i>Official Documents and Regulations</i>	31
Table 2. <i>Characteristics of the Participants (EFL Teachers) (continued)</i>	48
Table 3. <i>Linking Interview Questions for EFL Teachers to the Research Variables</i>	53
Table 4. <i>Linking Interview Questions for Field Experts to the Research Variables</i>	55
Table 5. <i>Interviewing Duration</i>	57
Table 6. <i>Cases, Codes, and Categories (continued)</i>	60

LIST OF FIGURES

<i>Figure 1. Cases and units of analysis in the current research study</i>	44
<i>Figure 2. Regions and cities in which EFL teachers work.....</i>	47
<i>Figure 3. Distribution of the universities at which the field experts work</i>	50
<i>Figure 4. The development process of the semi-structured interview form</i>	54
<i>Figure 5. Framework for teacher-based assessment at primary state schools.....</i>	67
<i>Figure 6. Washback effects and triggering factors in primary education</i>	72
<i>Figure 7. Use of mock tests at primary schools: External forces that mandate the use of mock tests and outcomes for language learning.....</i>	75
<i>Figure 8. Influential factors that increase the demand on the use of multiple-choice items included in language tests.....</i>	77
<i>Figure 9. Teacher-related factors that influence the intensity of washback effects</i>	80
<i>Figure 10. Washback effects of teacher-based assessment on EFL teaching dynamics and triggering force that increases exam anxiety</i>	83
<i>Figure 11. Use of mock exams in lower secondary education, force triggering their use and washback on language learning and testing</i>	85
<i>Figure 12. Language teaching shift in lower secondary education</i>	88
<i>Figure 13. Radical shift between the grades in lower secondary education in terms of English language testing</i>	94
<i>Figure 14. Institutional policies that are developed for USE-EE preparations</i>	97
<i>Figure 15. Washback effects of USE-EE on language teaching, testing and policy-making</i>	98
<i>Figure 16. The relationship between exam-oriented language teaching and influential factors that trigger exam-oriented language teaching.....</i>	102
<i>Figure 17. Operationalization of USE-EE washback effects triggered by macro- and micro-level dynamics in lower secondary education</i>	109

<i>Figure 18. Washback effects in upper secondary education and main triggers that exert the effects of teacher-based formal examination.....</i>	<i>117</i>
<i>Figure 19. Washback effects of FLT and BCT and language teaching, learning, testing, learners and teachers</i>	<i>127</i>
<i>Figure 20. English language teaching and testing practices in upper secondary education: Radical shift from the grade 9-10 to the grades 11-12.....</i>	<i>128</i>
<i>Figure 21. LASBLA: Pyramid of large-scale skills-based language assessment.....</i>	<i>138</i>
<i>Figure 22. The triggering factors that bring about the effects of centralized large-scale examination on EFL teaching dynamics</i>	<i>142</i>



LIST OF ABBREVIATIONS

MoNE	Ministry of National Education
CEFR	Common European Framework of References
CoE	Council of Europe
CoHE	Council of Higher Education
USE-EE	Upper Secondary Education Entrance Exam
BCT	Basic Competency Test
FCT	Field Competency Test
FLT	Foreign Language Test
ELF	English as a Foreign Language
MSCP	Measurement, Selection and Placement Center
SEM	Structural Equation Modeling

CHAPTER I

INTRODUCTION

This chapter includes the major concepts and considerations of the washback in language testing which is the focus of this research. It begins with the definitions and terminology of the washback. In what follows, the basic types of washback effects are accentuated, and the social, educational, and political dimensions of language testing are discussed with reference to the consequences of language test use. With regard to the factors influential in bringing about washback effects, teacher-related variables are elaborated in detail. In connection with these theoretical bases, the problem, aims, significance, assumptions and limitations of the study are explained, and finally the basic terminology is provided.

Problem Statement

As a crucial part of language teaching, English language testing is a comprehensive field which encompasses a wide range of components. Language tests, which are regarded as the basic component of language testing, are the vehicles used extensively to measure test takers' performance in the target construct. Bachman and Palmer (1996) describe the main qualities of a language test as reliability, construct validity, authenticity, interactiveness, impact, and practicality. In this description, 'impact' refers to the effect of language tests on the dynamics of language teaching and learning. Terminologically speaking, the effect of language testing on language teaching is called *washback effect* (Brown & Hudson, 1998; Hughes, 2003; McNamara, 2000). In the relevant literature, there are various terms which are coined to specify the washback effect such as *measurement-driven instruction* (Popham, 1987), *systematic validity* (Frederiksen & Collins, 1989), and *consequential aspect of the construct validity* (Messick, 1996).

According to Brown and Hudson (1998), there are two main types of washback effect: positive and negative washback effects. Whereas the positive washback emerges when “the assessment procedures correspond to the course goals and objectives”, the washback becomes negative “if the assessment procedures do not correspond to a curriculum’s goals and objectives” (Brown, & Hudson, 1998, p. 667). Messick (1996) mentions the minimization of “the sources of invalidity” to promote positive washback. These invalid sources are regarded as “construct underrepresentation” and “construct-irrelevant variance” which are respectively linked to the authenticity and directness of language tests (p. 4). Bachman and Palmer (1996) claim that “the impact of test use operates at two levels: a micro level, in terms of the individuals who are affected by the particular test, and a macro level, in terms of the educational system or society” (p. 29). Supporting this explanation, Cheng, Sun, and Ma (2015) make a basic distinction among the terms “washback”, “consequence”, and “impact”. According to this distinction, *washback*, at micro level, refers to the effect of language testing on language teaching; *consequence* is directly related to “the results of the (mis)use of test scores” (p. 437); and *impact*, at macro level, indicates “the extended influence of testing beyond classroom teaching and learning to other stakeholders in education” (p. 438).

Language test results, which provide a basis for the instructional decisions, have important consequences for all stakes (Shohamy, 1993; Shohamy, Donitsa-Schmidt, & Ferman, 1996). These decisions may be exemplified as the placement to a program, completion of a course and receiving a certificate, application for a job, curriculum and syllabus design, material selection, and educational policies (Shohamy, 1993; Taylor, 2005). As seen from these consequences, language tests affect social, educational and political decisions in the EFL (English as a Foreign Language) teaching. McNamara (2000) states that “language testing, like language itself, cannot ultimately be isolated from wider social and political implications” (p. 77). In relation to the use of language testing at societal level, Fulcher (2009) calls attention to the case that the use of language tests and interpretations of test scores may change according to the nature of the society. Along with this societal impact of language testing, Cumming (2009) asserts that language testing itself is “a major policy issue in educational programs” (p. 91). Through the introduction of the Common European Framework of Reference (CEFR) and declaration of the Bologna process (Fulcher, 2009), the global standardization has become prominent in the educational settings, and this

tendency has affected the development and use of language tests, and consequently language testing policies. Reviewing the research on the washback effect, Cheng, Sun, and Ma (2015) categorized the studies according to the research themes, one of which argues that language testing is recognized as “a policy instrument to promote language education.” (p. 439). For example, in one of the preceding studies (Wall & Alderson, 1993), language examination was used to “to reinforce the changes suggested by the textbooks” (p. 4). As Shohamy (2007) supports, “language tests, given their power and influence in societies, play a major role in the implementation and introduction of language policies.” (p. 120). According to Ross (2011), language testing is currently regarded as “a political act” and associated with “the issues of power, identity, national sovereignty, macro- and micro-politics, as well as macro- and micro-economics” (pp. 786-795). Considering the power of language tests, Shohamy (1998) puts forward critical *language testing* and defines it as “both a product and agent of cultural, social, political, educational, and ideological agendas that shape the lives of individual participants, teachers and learners” (p. 332). In light of the considerations emphasized in the critical language testing, she explicates how a test becomes a powerful and critical tool in a society through a concrete example. That is to say, language tests, which hold the powers of the social and educational dynamics, are widely utilized as a tool to launch or implement new instructional practices (Shohamy, 1993).

In language testing, washback effect, acting on the social, practical and political stages, is a complex phenomenon and influenced by a large number of variables (Shih, 2009; Shohamy et al., 1996; Spratt; 2005; Watanable, 2004). These variables, which interplay with the washback, are holistically identified as decision-making agencies, intervening agencies, and implementing agencies (Cheng, 1999). As the major agent of the implementation, language teachers play a central role in how washback operates in the instructional settings. Drawing on the findings of the seminal research studies, Spratt (2005) echoes the teacher-related factors influential in bringing about washback effects such as teacher beliefs, attitudes, and education. In some research studies (Chen, 2002; Cheng, 1997; Wall & Alderson, 1993), there is a main emphasis on the finding that washback may have an impact on the content of teaching, but teachers’ methodology does not change radically. In this respect, teachers’ beliefs about language teaching and learning gain importance (Wall & Alderson, 1993; Cheng, 1999). In addition to their beliefs on teaching, their knowledge, skills, principles and beliefs about language testing authorize how washback effects are exerted and manipulated

in the classroom. Winke (2011) explains that “because they have personal knowledge of the students and come to directly understand how testing affects them in their day-to-day lives, teachers are well positioned to recognize discrepancies between classroom and test practices”; therefore, their perceptions on washback on teaching are significant (p. 633). From political aspects, the relevant literature suggests that further studies should be carried out at micro- and macro-levels to “explore the relationship between washback and other forces operative in the educational context” by taking into consideration “social, political, and economic factors governing teaching and learning in relation to a testing program” (Cheng et al., 2015, p. 463). In addition, teacher assessment should be investigated to sharpen the understanding of “how policies and mandates mediate teachers’ practice and interests in language assessment” (Yan, Zhang, & Fan, 2018, p. 167). Agreeably, Menken (2017) recommends the researchers to study “how testing shapes and affects language education policies” at macro levels (p. 394). “Understanding washback requires a careful examination of the complex socio-political context in which the teaching-testing relationship is embedded” because such contextual factors play an essential role “shaping the relationship between teaching and testing.” (Ali & Hamid, 2020, p. 142). In light of these research perspectives, this study attempts to find out the macro- and micro-level operative factors exerting washback effects in terms of the socio-economic and socio-political consequences. Besides, language teachers’ instructional and assessment practices and beliefs will be examined in order to explicate how washback effects manifest themselves in the instructional settings and recognize the reflection of the macro language testing policies on the micro-level classroom practices.

Aim of the Study

Language testing has significant consequences for the stakeholders, who consist of “test-takers, teachers, test developers, administrators, community agencies, public officials, and researchers” (Norton, 1997, p. 313), in the socio-economic and socio-political contexts beyond its psychometric properties such as reliability, validity and objectivity (Shohamy, 1993; Bachman, 2005). In consideration of the social, economic and political dimensions of washback effects, Cheng et al. (2015) suggest that washback and operative forces should be researched, and the relationships between them should be highlighted. Likewise, Yan et al. (2018) encourage others to conduct a study to understand how teacher assessment practices

are governed by large-scale policies. With regard to the Turkish context, it is obvious that the national educational system includes a wide range of large-scale and high-stakes centralized language exams which perform as the gatekeepers for better educational opportunities (Alptekin & Tatar, 2011; Özmen, 2011a). Evidently, these exams exert numerous washback effects on the classroom teaching, educational outcomes, socio-economic decisions and language teaching and testing policies. Accordingly, this study aims to investigate the macro-level and micro-level triggers that are influential in how the washback effect of low-stakes and high-stakes assessment operates in EFL teaching and learning in Turkey. At micro level, teacher-based assessment is examined in order to observe the reflection of language teaching and testing policies on action and also provide insights into how washback manifests itself in low-stakes assessment. Following the familiarization with the classroom practices, the underlying forces of washback effects on EFL teaching are elaborated within nation-wide policies. To summarize, this study intends to explore how washback effects function in primary and secondary language education in Turkey and then find out external and internal triggers that shape how these effects operate in EFL teaching dynamics.

Research Questions

Considering the research aims, this study evolves around two main constructs: the washback effects of language testing and factors triggering the effects. Accordingly, the following research questions are taken into account in the current study:

1. How does language testing influence EFL teaching dynamics at the state schools in Turkey?
2. What triggers the washback effects of language testing on EFL teaching dynamics at the state schools?

The concept of language testing here refers to both low-stakes and high-stakes assessments that are administered at state schools in the MoNE. EFL teaching dynamics embraces teaching and testing methodology, language content and materials, and the stakeholders such as students, parents, teachers, and administrators. Sub-questions and research settings are comprehensively presented in the chapter of Methodology. These two main research questions function as a guide in reviewing literature, gathering qualitative data, analyzing

the obtained data and displaying the findings. This research focuses on only two parts (i.e. primary and secondary education)¹ of formal education at the state schools. The rationale why pre-school education and higher education are not included in the study is that there is no specific curricular background to teach English in pre-school education, and higher education institutions are autonomous to determine their own English language teaching and testing policies, practices and procedures. Hence, primary, lower secondary and upper secondary education levels are thoroughly examined with respect to the research constructs: washback effects and triggering forces.

Significance of the Study

The research on washback effects of language testing basically shows that language exams influence the dynamics of EFL teaching and learning to varying degrees. In some studies, the high-stakes or centralized language tests were extensively examined in different countries to point out how and to what extent these exams influenced the instruction (Andrew, Fullilove, & Wong, 2002; Chen, 2002; Cheng, 1997; Cheng, 1999; Cheng, 2003; Choi, 2008; Green, 2006; Green, 2007; Hung, 2012; Luxia, 2007; Muñoz & Álvarez, 2010; Shih, 2009; Shohamy et al., 1996; Wall & Alderson, 1993). These studies chiefly elucidate the impacts of language exams on the specification of the course content, selection of language teaching activities and materials, and learning outcomes. In a similar vein, the washback research in Turkey mainly focuses on the educational, social and economic effects of large-scale foreign language exams on the stakeholders (Akpınar & Cakildere, 2013; Çakır, 2017; Göktürk-Sağlam, 2018; Hatipoğlu, 2016; Kılıçkaya, 2016; Külekçi, 2016; Özmen, 2011a; Özmen, 2011b; Yıldırım, 2010). On the other hand, there has been limited research on how national educational policies on foreign language testing shape the nationwide operations of the washback effect, and how these policies are shaped by the washback effect itself at macro level. What's more, most research studies highlight teachers' beliefs, training and attitudes towards testing which may trigger washback effects in language learning. The current study will contribute to the field of English language testing not only by presenting the social, political, economic and educational dynamics of washback effects in Turkey but also by illustrating the role of EFL teachers' testing beliefs and

¹ Primary Education (Grade 1 to 4), Lower Secondary Education (Grade 5 to 8) and Upper Secondary Education (Grade 9 to 12)

practices in mediating the washback effects in the instructional contexts. In addition, most washback studies evolve around global or local high-stakes tests. This research also investigates how washback mechanisms of low-stakes assessment function in primary and secondary education, and how they are related to high-stakes washback triggers and outcomes.

Assumptions of the Study

It is noteworthy to express some important assumptions of this study. During data collection, semi-structured interviews were employed, and these interviews attempted to elicit in-class teaching and assessment practices. It is assumed that these responses correspond to their practices in real classroom settings that could not be observed due to the outbreak of the pandemic. EFL teachers were sampled from seven different regions and twenty-three cities. It is supposed that teacher responses represent the beliefs and practices of all EFL teachers who currently work in MoNE. Before conducting interviews, the interviewer asked all the participants about their volunteer contributions to the study and confirmation about informed consent form. Considering this process, it is believed that all participants took part in the interviews willingly and answered all the questions sincerely. Finally, washback, impact and consequences are the terms that may cover different circles of test effects (Cheng et al., 2015). This study mainly used the concept of washback effect to represent the effects on “teaching and learning, the educational system, and the various stakeholders in the education” (Andrews, 2004, p. 37).

Limitations of the Study

The focus of this study is relatively large-scale and multi-dimensional; for this reason, the cases of the study need to be examined from different lenses. However, due to the outbreak of Covid-19 pandemic, qualitative data were gathered only through the interviews. Whereas semi-structured interview is the only way of data collection in this research, it has been consolidated with some techniques. For example, maximum variation sampling was used to choose the teacher participants. That is, EFL teachers were selected with reference to their educational background, place of work and teaching experiences in order to mirror how washback effects function in three levels of education. Since the classroom setting was not observed and other stakeholders except for field experts were not interviewed,

comprehensive questions were directed to both language teachers and experts in relation to the units of analysis. A cross-sectional research perspective was adopted in this research; namely, data collection was conducted only for one academic year. Longitudinal studies may give further insights into how washback effects and macro- and micro-level triggers change in time. On the other hand, this study collects data from the grades 2 to 12. In other words, it is attempted to observe the longitudinal influence of language exams within a cross-sectional study. Therefore, how washback effects differ in transition to upper or lower grades are presented in consideration of EFL teaching and testing dynamics.

Definitions

Assessment Literacy: It is “individual’s understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions” (Popham, 2011, p. 267).

Consequential Validity: It is a kind of validity framed by Messick (1996) and underscores that “the social consequences of test use and interpretation support the intended testing purpose(s) and are consistent with other social values.” (Shaw & Weir, 2007).

High-stakes Testing: It refers to the importance of a test when “a single test score is used as the main or sole factor in determining significant educational decisions” (Menken, 2017, p. 386).

Low-stakes Assessment: This kind of assessment has outcomes that are not so influential in decision-making about the stakeholders (Fulcher, 2013).

Measurement: It is regarded as “the process of quantifying the characteristics of persons according to explicit procedures and rules” (Bachman, 1990, p. 18).

Mock Examination: They are sample forms of large-scale tests that are taken by individuals in order to practice large-scale implementations and make preparations for the real administration in advance.

Stakeholders: They are generally “learners, teachers, test developers, administrators, community agencies, public officials, and researchers” who are those holding a stake in testing (Norton and Stewart, 1999, p. 224).

Teacher-based Assessment: It is namely classroom-based assessment that includes teacher-led practices that gather “multiple types of information concerning student language use, to analyze and interpret it, to provide feedback, and to use this information to help make decisions to enhance teaching and learning.” (Turner, 2012, p. 65)

Test Blueprints: “A blueprint consists of characteristics pertaining to the structure, or overall organization, of the test, along with test task specifications for each task type to be included in the tests.” (Bachman & Palmer, 1996, p. 90). Tests are basically “constructed according to precise specifications, a ‘blueprint’ outlining a procedure for test development (Douglas, 2010, p. 6).

Test Impact: The influence of language tests on teaching dynamics beyond the classroom such as educational system, society or policies (Taylor, 2005)

Washback: “The effect of a test on teaching and learning, including effects on all aspects of the curriculum, including materials and teaching approaches, as well as on what the students do to learn and to prepare for tests.” (Carr, 2011, p. 332)



CHAPTER II

REVIEW OF LITERATURE

This section firstly presents the main definitions and concepts related to washback in addition to washback dimensions, models and hypotheses. It also describes the main research studies which were carried out for the purposes of elucidating the washback effects of language tests or examinations on language teaching and learning. The washback studies are reviewed here in two categories: the international contexts and the Turkish context in which mostly high-stakes and large-scale language tests were examined in terms of the stakeholders, teaching methodology, test specifications, classroom practices, learning improvement, and socio-political aspects. Finally, English language teaching and testing policies are historically explicated in regard to the primary and secondary education.

Basic Concepts in Language Testing

Language testing is a comprehensive and far-reaching field in which a wide range of concepts are used to imply the assignment of the values to the operationalization of target traits and make judgments on these traits. These concepts are basically known as *test*, *measurement*, *assessment*, *testing* and *evaluation*.

Test is viewed as “a measurement instrument designed to elicit a specific sample of an individual’s behavior” (Bahman, 1990, p. 20). Brown (2004) specifically indicates that tests are the tools that measure test-takers’ ability, knowledge or performance in the target field. These tools are “the range of alternatives available to language educators for gathering trustworthy information – information that is then used to inform various interpretations and actions.” (Brown, Hudson, Norris, & Bonk, 2002, p. 13). *Testing* is the act of employing

tests that involve “a set of techniques, procedures, or items” and require test-takers’ performance to be measured (Brown, 2004, p. 3). They are principally used to generate benchmarks to make sensible interpretations (Hughes, 2003). In this way, they provide information for evaluative decisions.

Measurement is mainly based on the quantification in which numerical values are attributed to test takers’ performance (Davies et al., 1999). Concerning language testing field, it deals with quantifying language-related outputs. As a broader concept, *assessment* is seen as “the process of collecting information about a given object of interest according to procedures that are systematic and substantively grounded” (Bachman, 2004, p. 7). According to Purpura (2016), it is the systematic elicitation of the data gathered through tests or other instruments in order to provide a basis for making inferences or decisions on individuals’ language performance. Finally, *evaluation* is a set of procedures based on making decisions on what is assessed, tested or measured (Bachman, 1990). It draws on the data gathered from assessment or testing to make judgments on the language characteristics. Genesee and Upshur (1996) call attention to three aspects of evaluation: information, interpretation and decision-making. Information is obtained through measurement, assessment or testing. Interpretations are proceeded to make sense of the available information; in doing so, “it becomes meaningful” (p. 4) Finally, making-decision is performed in consideration of the interpreted information.

Taken together, a number of terms are used in the field of language testing to make obvious the procedures of collecting information, quantification, interpretations and making decisions. Bachman (2004) visualizes the relationships among these terms. Accordingly, test is the smallest unit operated at the core of the framework because it supplies the data for measurement, assessment and evaluation. Assessment plays a role in this framework as an umbrella term that covers the concepts of measurement and test since it is a continuous information-gathering process through tests or other tools. In the scholarly papers or academic publications, language testing and assessment may be used interchangeably but it is noteworthy to state that testing is “one form of assessment” (Hughes, 2003, p. 5). Evaluation is mostly interfaced with assessment, measurement and tests in terms of utilizing the information collected. However, in evaluation, there is a separate mechanism that compares the performance with pre-determined benchmarks and makes judgments on to what extent this performance meets the requirements of these benchmarks.

Test Purposes and Classifications

Language tests can be categorized in different ways according to various classification criteria. McNamara (2000) uses test method and purpose as the criterion point to make a distinction among different types of tests. Hughes (2003) pays attention to the information provided by tests in his classification and discusses four different types of tests: proficiency, achievement, diagnostic and placement tests. Similarly, Brown (2005) points out the decisions that are made on the basis of test results. Carr's (2011) classification is grounded in Brown's (1995) decision-based categorization, and it is mainly divided into two sections: curriculum-related decisions and other decision types. Admission, placement, diagnostic, progress and achievement tests are grouped into the category of *curriculum-related decisions* while proficiency and screening tests are classified in *other types of decisions*. This classification, test types and purposes are elucidated below in relation to their role in decision-making.

More particularly, *admission tests* are employed at the beginning of a specific program in order to make decisions on the individuals who will get training (Carr, 2011). *Placement tests* hold "the goal of grouping students of similar ability levels together" (Brown, 2005, p. 9). When the individuals are selected into specific programs, placement tests are conducted to bring the students with similar abilities together with the aim of more homogenous classes. *Diagnostic tests* are "designed to diagnose specified aspects of a language" (Brown, 2004, p. 46). They are primarily used to recognize whether individuals have target characteristics of the language. When the individuals are admitted to a specific program and placed into the classes or grades in accordance with their ability levels, diagnostic tests can be carried out to measure the individuals' readiness in course content. When the course procedures are initiated and individuals practice learning objectives, progress tests can be presented to get familiarized with learning strengths and weaknesses. That is, *progress tests* are administered to keep track of individuals' ongoing learning process in relation to course objectives (Carr, 2011). Finally, during or at the end of the course, *achievement tests* are employed to define the level of attainment. They put evidence on "the extent to which an individual has mastered the specific skills or body of information acquired in a formal learning situation" (Harris, 1969, pp. 3-4). Conceptually, the abovementioned tests provide information for curriculum-related decisions on the individuals like admitting, placing, recognizing, monitoring and determining the attainment.

In the second category, proficiency and screening tests are utilized to give insights into other decisions. For example, *proficiency tests* “can measure overall mastery of English or how well prepared one is to use English in a particular setting” (Madsen, 1983, p. 9). In this respect, test results can be interpreted in different contexts in order to show the individuals’ level of English. According to Hughes (2003), “the content of a proficiency test, therefore, is not based on the content or objectives of language courses”; instead, “it is based on a specification of what candidates have to be able to do in the language” (p. 11). This illustrates the basic distinction between achievement and proficiency tests. Whereas the achievement tests gather information about the individual’s attainment levels in a specific language program, proficiency tests are independent of the any curricular objectives and constructed with regard to any language’s mastery levels.

As another type of test taken into account in making other decisions, *screening tests* are used to examine “whether someone is sufficiently proficient in the target language to be qualified for a particular job”; in other words, this kind of testing is the utilization of proficiency tests for selecting purposes (Carr, 2011, p. 9). It is also highlighted that there are some overlapping points between screening tests and admission tests (ibid., 2011). Screening tests can be administered for educational purposes; from this perspective, they may function in admission testing. Besides that, screening tests can be used for non-educational purposes like a selection of a job candidate. That is why they also could be categorized in *other types of decisions*.

Qualities of Language Tests

Bachman and Palmer (1996) introduced the notion of test usefulness by depicting the basic qualities of language tests. According to their approach, there are six main qualities that “determine the overall usefulness of a given test” (p. 19): reliability, construct validity, authenticity, interactiveness, impact and practicality. While the first two qualities are related to measurement aspects of language tests, others may be attributable to test relations with language instruction, learning tasks and practices.

More specifically, *reliability* is the measurement consistency of language tests. “A reliable test will be consistent across different characteristics of the testing situation.” (ibid., p. 19). In a similar vein, Douglas (2010) notes that reliability deals with consistent measures on

different occasions. The aspect that a language test presents alike results in various contexts contributes to test usefulness. Construct validity is another measurement-related quality that indicates to what extent test results are linked to the characteristics that are intended to measure (Brown, 2005). In order to enhance this quality, it is important to define test constructs operationally “in a manner that can serve as a basis for score interpretation” (Chapelle, 2012, p. 27). Bachman and Palmer (1996) raise awareness on the importance of the construct validity in the interpretations of test scores. For that reason, Weir (2005) views their validity approach as “a posteriori relationship between a test and psychological abilities” (p. 17). It can be concluded that construct-based interpretations of test scores promote test usefulness.

In addition to measurement-focused qualities, *authenticity* is concerned with the connection between test specifications and target language use. In other words, it is “the degree of correspondence of the characteristics of a given language test task to features of a TLU task”. (Bachman & Palmer, 1996, p. 23). Brown (2004) presents that authenticity can be ensured through the use of natural language, contextualized and thematically organized language items, relevant and interesting assessment topics, and real-life assessment tasks. As for *interactiveness*, this quality basically refers to the interaction between test-takers and test itself. Bachman and Palmer (1996) define it “as the extent and type of involvement of the test taker’s characteristics in accomplishing a test task” (p. 25), and they exemplify these characteristics as language knowledge, topical knowledge, metacognitive strategies and affective schemata. In other words, test-takers utilize their background knowledge in relation to test topics as well as language ability, strategy and schemata. As another test quality, *practicality* deals with the resources needed to develop, administer and score language tests and interpret test results (Brown, 2004; Brown, 2005). Bachman and Palmer (1996) explain that practicality is “a matter of the extent to which the demands of the particular test specifications can be met within the limits of existing resources” (p. 36) and shed light upon three types of resources: human resources (item developers, raters, testers), material resources (place, technological equipment, materials) and time (development, administration, scoring). Test usefulness also depends on the cost-effective use of these human-, material-, and time-related resources.

Finally, impact is known as the test aspect that is directly associated with the use and interpretations of test scores in a school community or a wider setting like education system

and society. In literature, different studies are in attempt to give the definitions related to impact, washback, and consequences (Alderson & Wall, 1993; Cheng, Su & Ma, 2015; Hamp-Lyons, 1997; Taylor, 2005). Hamp-Lyons (1997) discusses the concept of impact at macro- and micro-levels. When test impact influences micro-level dynamics like classroom teaching, course content, or learning, it is called “washback effect”. More particularly, *washback* is the effect of language tests on language teaching (Alderson & Wall, 1993). On the other side, *impact* refers to the influence of language testing beyond the classroom. It could be on educational system or society (Hamp-Lyons, 1997). As a more extensive term, *consequences* are related to the use or misuse of test outcomes (Cheng, Sun, & Ma, 2015). Accordingly, a language test can be useful when it brings about intended impacts on teaching dynamics or broader settings as determined in test policies and specifications.

Washback

Washback, in simple terms, is defined as the effect of language testing on teaching. It is also interchangeably used with the term “backwash” (Hughes, 2003). Cheng and Curtis (2004) provide the examples of dictionary meanings of “backwash” and express negative connotations of the definitions. In applied linguistics, these two terms correspond to both positive and negative effects of testing. Brown (2006) holds an educational point of view and defines washback as “the degree to which a test affects the curriculum that is related to it” (p. 242)

Historically, washback has been used with different terminology: measurement-driven instruction (Popham, 1987); systematic validity (Frederiksen & Collins, 1989); test-curriculum alignment (Shepard, 1990), and consequential validity (Messick, 1996). More specifically, Popham (1987) states that “measurement-driven instruction occurs when a high-stakes test of educational achievement, because of the important contingencies associated with the students’ performance, influences the instructional program that prepares students for the test.” (p. 680). This article also discusses the role of measurement-driven instruction in the improvement of educational quality. Frederiksen and Collins (1989) introduce systematically valid tests into educational systems and draw attention to facilitating influence of these tests on “the development of cognitive traits” to be measured (p. 27). Test-curriculum alignment is another term that associates with washback effects, and

Shepard (1990) mentions the intended act of aligning teaching programs with tests within this terminological framework. Similar to the concern of validity in Frederiksen and Collins (1989), Messick (1996) brings consequential validity into the spotlight by discussing minimization of invalidity sources. He underlines that construct underrepresentation and construct-irrelevant variance are two threats to validity; so, the minimization of these factors fosters positive washback.

Basically, there are two types of washback: positive and negative (Brown & Hudson, 1998). If testing is designed in line with the curriculum and supports the implementation of the curricular content and methodology, it brings about positive washback. Otherwise, when it is not in tandem with teaching rationale and causes narrowing of curriculum, it means that testing exerts negative washback. Broadening the scope of washback, Watanable (2004) presents five dimensions of washback: specificity, intensity, length, intentionality, and value. These dimensions are scrutinized as follows (pp. 20-21):

- *Specificity* views the washback from general or specific parameters. General washback is generated by any kind of test whereas specific washback is triggered by a particular aspect of a test.
- *Intensity* is the washback variable coined by Cheng (1997). In this dimension, intensity could be strong or weak. Depending on the degree of effect, instructional practices are partially or largely influenced by tests or examinations.
- *Intentionality* covers intended and unintended washback effects. In test specifications or development, some intended consequences may be considered. On the contrary, there could be some unexpected or unintended outcomes of tests.
- *Length* indicates the duration of test effects. Some of them could be long-lasting; on the other hand, some washback effects may be short-term.
- *Value* is typically categorized as positive or negative effect. In consonance with the discussion about washback types above, intended washback may be linked to positive washback.

The relevant literature lists some guidelines, insights and strategies that help promoting positive washback. Madsen (1983) states that language testing may give the students “a sense of achievement”, and it may function as the source of feedback that displays teaching or learning strengths and weaknesses (p. 4). Hughes (2003) recommends the stakeholders to

assess what they expect to improve, ensure content validity, employ direct assessment, consider criterion-referenced perspectives, take into account language objectives, make the test understandable for all parties, “where necessary, provide assistance to teachers”, and “counting the cost” (pp. 53-56). In a similar vein, Bailey (1996) presents some insights about beneficial washback in the seminal work. Accordingly, all the significant others should be informed about test purpose; test results should be “believable”, “credible” and “fair”; construct validity should be ensured; testing should be built on predetermined objectives and “some theoretical principles”; tests should include authenticity; and tests should get the learners involved in the assessment (p. 275-276).

As a landmark study in washback field, Alderson and Wall (1993) formulate fifteen washback hypotheses that frame possible effects of testing on teaching, learning, what to teach and learn, attitudes, and stakeholders. Accordingly, a test will have an impact on teaching, learning, teacher and learner methodology, course content, “the rate and sequence” of learning and teaching, “the degree and depth” of learning and teaching, “attitudes towards the content, method of learning and teaching”; “tests that have important consequences will have washback”; “tests that do not have important consequences will not have washback; tests will influence “learners and teachers” or “some learners and teachers” (pp. 9-10). Presumably, these hypotheses outline the variables that are likely to be influenced by tests. Along with washback hypotheses, some mechanisms and models are suggested in the washback field. Bailey (1996) develops a washback model by drawing on Hughes’ trichotomy and Alderson and Wall’s (1993) fifteen hypotheses. In this model, there are three major parts: participants, processes and products that all are connected to “test”. Also, model constructs are in relationship with others in different ways. In 1999, Cheng identifies three agencies: decision-making agencies, intervening agencies, and implementing agencies in her “Washback explanatory model” (p. 257). Another different washback model is proposed by Shih (2009) who places washback on teaching in the center of the model and builds links to this focal construct from contextual factors, test factors and teacher factors. In this model, all micro- and macro-dynamics and factors that exert washback effects or influence operationalization of washback are demonstrated. It can be deduced from the hypotheses and models that a number of variables perform in washback mechanism. As a complex phenomenon (Watanabe, 2004), washback has been researched in both international and local settings. The following chapters summarize the washback-related studies.

Research Studies on Washback in International Contexts

As considered one of the starting points in the field of the washback research, Wall and Alderson's study (1993) threw light upon the impact of the O-level test examination on English language teaching in the Sri Lankan secondary education. The trigger force to administer this exam was "to reinforce the changes suggested by the textbooks" (p. 4). The study findings uncovered some positive and negative washback effects of the O-level test on language teaching. Whereas "teachers' determination to cover the textbooks" was regarded as the positive aspect of washback (p. 27), the focus on the language skills only assessed in this exam affected instruction negatively. Also, it is seen that the content of teaching was influenced by the O-level testing to some extent but there was no absolute impact on teachers' how to teach English. This study also shows that it is essential to increase teacher knowledge and awareness on test demands. In her sequential studies, Cheng (1997; 1999; 2003) investigated the impacts of HKCEE (Hong Kong Certificate of Education Examination in English) in terms of the instructional practices, and teachers' perceptions and attitudes. This exam was introduced to the educational settings in order to exert positive washback effect on language teaching in secondary education. In the study conducted in 1997, she found that "teaching content received the most intensive washback effects" (p. 16), language teachers' choices of in-class practices and teaching materials were also affected by this examination but no changes were observed in their teaching methods. In 1999, Cheng drew attention to the washback effects on teachers' perceptions and practices. Similar to the previous study findings, it is illustrated that the modification in language testing might bring about changes in teaching content and activities as required in the new examination, but this modification influenced EFL teachers' major beliefs and practices to varying degrees. Teachers came up with new ideas and activities depending on their understanding of modification. In this study, Cheng also identified the agencies of the washback effect as "decision making agencies" (e.g. teaching and exam syllabus), "intervening agencies" (e.g. teaching content and methods) and "implementing agencies" (e.g. teachers and students) through her washback exploratory model (p. 257). In her another study on the effects of public examination change in Hong Kong (Cheng, 2003), she emphasized once again that language teaching activities in secondary education were shaped by the examination itself, but language teachers' methodology remained the same.

Shohamy, Donitsa-Schmidt, and Ferman (1996) focused on the impacts of two main language exams conducted in Israel. One of these exams was constructed to measure the learners' language performance in ASL (Arabic as a Second Language), and the other one was based on testing EFL (English as a Foreign Language) skills. After defining the washback effects of the exams separately, the researchers compared them in consideration of the socio-political dynamics of language education. The main results show that whereas slight modifications did not bring about any changes in instructional and testing practices in the ASL context, they caused large effects on teaching, test preparation, and teaching materials in the EFL context. The study indicates that "washback can change over time" and "it depends on a number of factors" which are identified as the stakes of the test, status of a language, test purpose, test content, and language skills tested (p. 314). In another educational context in Taiwan, Chen (2002) questioned the impacts of a newly introduced high-stakes public exam on English teaching, and teachers' perceptions about these impacts. The major findings of the study highlight the perspective of "teaching to the test" (p. 16). In other words, teachers arranged their classroom practices (e.g. language content, skills, and activities) in agreement with the requirements of the examination. More specifically, teachers conducted more speaking tasks because spoken production was initiated to measure in the new system. In this regard, Chen (2002) drew attention to providing language teachers with more professional development opportunities to adapt curricular and testing-related changes. Also, it is recommended that oral assessment should be placed into testing system. In connection with the influence of testing approaches on teaching practices, Yan et al. (2018), in their study attempting to conceptualize language assessment literacy and to define the assessment needs of language teachers, found that language teachers proceeded "the alignment between teaching and testing" (p. 165). Additionally, Chen's (2002) study points out that language testing affected the content of teaching, but it did not influence how to teach English.

In Hong Kong, Andrew, Fullilove, and Wong (2002) investigated the effect of a high-stakes exam, Hong Kong Advanced Supplementary "Use of English" Oral Examination, on English language learning. This exam influenced the learners' spoken performance to some extent, but "the impact of the test on student performance has been delayed" (p. 220). Also, it is stated that the second phase provided more evidence on washback effects; however, these effects indicated "a very superficial level of learning outcome: familiarization with the exam

format, and the rote learning of exam-specific strategies and formulaic phrases” (p. 220). They also mentioned the complexity of washback effects. In another washback study, Choi (2008) argued the effects of various high-stakes tests (e.g. TOEIC, TOEFL, and YLE) administered in EFL teaching in Korea at different levels of education. The study findings reveal some negative washback effects on learning. For example, it is stated that EFL exam made the pupils feel under pressure to take it, and the learners in secondary education needed to spend more time in order to get prepared for “college entrance exam” (p. 58). These high-stakes exams also generated the negative consequences about the financial issues.

Luxia (2007) aimed to describe the intended impacts of a nationwide high-stakes test, the NMET (The National Matriculation English Test), on EFL writing in China. While the expected outcomes were achieved to some extent, some aspects of test implementation needed to be addressed with more support and collaboration among the stakeholders. Specifically, the NMET affected the content and procedures of teaching as intended along with the due attention and time. On the other hand, language teachers had a tendency to keep teaching with their classical approaches. The study clarifies that the introduction of a high-stakes test to an education system is not sufficient alone to accomplish the test objectives; other influential factors should be taken into consideration. Shih’s (2009) study analyzed the impact of the General English Proficiency Test (GEPT) on language teaching in Taiwan and discovered “minor and teacher-specific” washback effects of this exam on teaching practices (p. 198). Shih (2009) claims that a wide range of factors influence “the degree of washback” (p. 198) and then presents a holistic picture of the factors contributing to the washback of tests. The main types of factors are illustrated as *contextual factors* (national, social, broader educational factors, school-level factors, and course-level), *test factors*, and *teacher factors* (p. 199). Also, this washback model displays the relationships among these major factor categories and explains what factors trigger the operationalization of washback. In a similar fashion, Spratt (2005) elaborates on the factors influential in the operation of the washback effects by documenting the relevant research studies in the field. The major emphasis was placed upon the teacher-related factors as teachers’ beliefs, attitudes, education and training, personality and willingness to innovate (p. 29). Following that, the resources, exam properties, and school factors were discussed in detail.

With respect to large-scale language exams recognized around the world, Green (2006; 2007) studied the washback effects of IELTS (International English Language Testing

System). In 2006, he examined two courses from teacher and learner perspectives: IELTS course, in which the focus was centered upon the test specifications of writing skill, and non-IELTS course, in which learners improved their overall writing skills. It is pointed out that language teachers in two different courses had “distinctive aims”, and learner perceptions were influenced by the course purposes and outcomes (p. 131). In another study, Green (2007) compared three groups in terms of the washback to IELTS learning outcomes. These three groups were categorized as (1) IELTS writing course, (2) pre-sessional EAP writing course, and (3) a course combining both IELTS preparation and EAP writing. The study findings showed that all students in three groups obtained almost the same level of attainment in writing, and this demonstrates that there is “no clear advantage for focused test preparation” (p. 1).

Muñoz and Álvarez (2010) examined the washback effects of Oral Assessment System (OAS) on the university-level speaking skills. This study was designed to compare the experiment group in which the learners participated in OAS and familiarized themselves with the oral aspects given in the rubric and the comparison group in which the learners did not attend OAS. The results explain that positive washback might be promoted through raising the learners’ awareness on the assessment processes, connecting assessment and instructional objectives and administering self-assessing practices. Regarding the alternatives in assessment, Hung’s (2012) study investigated the washback effect of e-portfolio assessment in EFL teacher preparation programs. The research findings indicate that this assessment method exerted both positive and negative washback effects. More particularly, “facilitating peer learning, building community of practice, enhancing learning of content knowledge, promoting professional development, and cultivating critical thinking” could be exhibited as the indicators for positive washback; on the other hand, “anxiety” and “frustration with technology” were among the negative effects of e-portfolio assessment.

Ali and Hamid (2020) explored the washback effects of SSC (Secondary School Certificate) exam on secondary education in Bangladesh. They interviewed with 12 teachers and supported the data with document analysis. The findings explain that “social value of grades, family expectations, institutional reputation” as social forces, “testing and curricular objectivity, task validity, score reliability” as assessment factors, and “hidden syllabus” contribute to the emergence of negative washback effects in Bangladesh. Exam results

“symbolize social and economic security for people-grades are social assets for parents, teachers and school’s prestige, students’ future academic and employment prospects, and security for schools in a competitive neoliberal market.” (p. 142). Zou and Xu’s study (2017) focused on program administrators as the stakeholders in order to familiarize with their perspectives on the influence of Test for English Majors for Grade Eight (TEM8) in China. They came up with both intended and unintended uses of test results. Program administrators have mostly positive attitudes towards the washback effects of TEM8 on course design, objectives, teaching and scheduling. However, washback intensity varied across instructional contexts. Tsagari (2011) examined the washback effects of the First Certificate in English (FCE) exam on teachers’ perceptions and practices. The basic findings indicate that there is a relationship between examination content and in-class practices; the exam had some negative affective influence on learners; because of high-stakes outcomes, FCE made teachers accountable for the results; and teachers’ negative perceptions affected their classroom practices. Xiao, Sharpling, and Liu (2011) investigated washback effects of National Matriculation English Test (NMET) on students’ learning performance. They noticed the impact on students’ learning strategies, particularly on reading skills. The students performed test practices and attempted to cover all the details in case they could be measured. The study findings also indicate the students’ use of metacognitive strategies.

Shih’s (2010) study focuses on the washback effects of General English Proficiency Test (GEPT) on the policies of universities in Taiwan. As a criterion for graduation, GEPT has some effects on tertiary-level policies that are triggered by social and educational, school, parental and student factors. Hence, teachers pay attention to these factors before their implementations. Zhan and Andrews (2014) examined the effects of College English Test Band 4 (CET-4) on non-English major students’ out-of-class learning experiences. They found out that learners’ beliefs about test characteristics like stakes, content, format and difficulty, self-knowledge, educational and test-taking background, significant others in school community, learning settings and possible CET-4 selves mediate the washback effects of CET-4 examination. Xie and Andrews (2012) used a statistical model, SEM (Structural Equation Model), to prove the relationships between mediating factors and washback effects. Their study findings explained that test design and use of high-stakes tests are influential in individuals’ test preparations. It is highlighted that test design has more extensive impact on these preparations. In 2015, Xie used SEM again to monitor mechanism

of washback and reached a conclusion that learners' perceptions brought about both positive (facilitating learning) and negative (test-oriented preparations) washback effects. In Wiyaka's study (2020), authentic performance-based reading test was examined in terms of its washback on learning and teaching in Indonesia. The research findings show that it promoted positive effects on students' attitudes and higher-order thinking skills. Language teaching was also influenced by methods, materials and allotted time. Drawing on a number of research studies, Tsagari and Cheng (2017) make a point that "washback is a highly complex phenomenon due to the fact that it is an interactive multidirectional process involving a constant interplay of varying degrees of complexity among the different washback components and participants" (p. 364).

Research Studies on Washback in Turkish Context

In the Turkish context, the research studies mainly describe the washback effects of high-stakes exams. From a holistic perspective, Toksöz and Kılıçkaya (2018) reviewed the research articles about the washback effects conducted in Turkey. They drew attention to the methodology and common findings which were mostly about the negative impacts of language tests on the instruction such as "teach-to-test" nature of language teaching, "narrowing of the curriculum in schools" and "anxiety" (p. 201). Özmen (2011a) presented a study on the washback effect of SEPPPO (The Selection Examination for Professional Posts in Public Organizations) as a large-scale exam used in Turkey in order to select personals for public institutions. He found that this exam had negative impacts on "allocated time for pre-service education", "financial issues", student teachers' motivation and exam validity itself. In his another study, Özmen (2011b) examined the washback effects of Inter-University Foreign Language Examination (ILE) on foreign language competences of candidate academics. The study findings pointed out that this exam had a number of washback effects on micro- and macro-level components of foreign language learning. More specifically, it had unintended effects on the candidate academics' "L2 competences, academic career, motivation, and cognitive learning" at macro level while it influenced the expenses for course materials, preparation time, and test quality negatively at micro level (p. 223).

Akpınar and Cakildere (2013) conducted a research study about the effects of high-stakes foreign language exams on academics' productive and receptive skills in the target language. They traced both positive and negative influences exerted on language skills. For instance, the negative aspect is pointed out in a way that the participants did not give much attention to the skills which were not tested in those exams. The findings show that there were statistically significant differences between the receptive and productive skills in terms of language development. Similar to the focus of this study, Külekçi (2016) examined the washback effects of foreign language examination employed widely in Turkey to measure the candidates' foreign language proficiency. The positive and negative effects are highlighted in the findings. For example, it provided learners with opportunities to practice in English and get promoted in a profession; on the other hand, the learners paid pivotal attention to the content and skills tested only in this examination. Also, it is stated that this exam affected the emergence of a new marketing area including test courses and practice books. Polat (2020) examined the large-scale language tests, Higher Education Foreign Language Test (YOKDIL) and Foreign Language Proficiency Test (YDS) that are developed and administered by Turkish agencies. The results indicate that test-takers are not able to get mastery in speaking, writing and listening skills that are not measured in these tests.

Yıldırım (2010) shed light upon the effects of a nationwide large-scale English language exam employed to place the candidates to foreign language-related departments at tertiary level in Turkey. He interviewed with the first year students in the program of English language teacher education. The study findings made explicit the negative impacts of the exam on the students' proficiency in English, prospective performance in pre-service teacher education, their "social skills and psychological well-being" (p. 113). In another study conducted at the university level, Göktürk-Sağlam (2018) argued the impact of University English Language Proficiency Test on teaching methodology in the EAP context in Turkey. She concluded that the proficiency test had a positive impact on teaching materials but influenced the course content and instructional practices negatively by narrowing down the scope of the curriculum through the aligning it with testing. Unlike the study findings which assert that language tests change course content but not methodology (Wall & Alderson, 1993; Cheng, 1997; Chen, 2002; Cheng, 2003), this research claimed that the washback effects of language testing are traceable in both teaching content and methodology at the university level of the EAP teaching.

Regarding high-stakes tests in lower secondary education, Kılıçkaya (2016) studied the washback effect of a centralized examination called TEOG (Transition Examination from Primary to Secondary Education). The study implied that teaching setting was heavily affected by the test requirements. According to the findings, this examination influenced the selection of teaching materials, “medium of instruction”, and skills to be taught, and it caused anxiety among the examinees (p. 131). Çakır (2017) carried out a similar study to elucidate the effects of the same high-stakes exam on the stakeholders. The study findings illustrate that teachers allocated much time to prepare the learners for this large scale exam. Also, it is pointed out in these findings that the stakeholders were influential in the specifications of teachers’ role in language testing. Uzun and Kılıçkaya (2020) examined the centralized exam Transition Examination from Primary to Secondary Education (TEOG) administered between 2016 and 2017 in terms of content validity. Accordingly, it seems that this examination has content validity in spite of some factors; and there is an alignment of the tests with course books. Considering the young learner context, Şenel and Tütüniş’s (2011) study presented that there was a negative effect of EFL writing assessment on young language learners’ performance especially when the pupils understood that their writing performances were being assessed since their level of anxiety increased during the assessment process.

Hatipoğlu (2016) focused attention on pre-service teachers’ perceptions on the impact of university entrance exam along with the historical movement of testing, selection and placement in Turkey. She carried out document analysis including a variety of official papers and conducted surveys and focus group interviews to elicit pre-service teachers’ perspectives on washback effects. The major findings explain that testing has gone through a number of shifts throughout the history. University entrance examination basically causes teaching-to-test practices, exam-led hidden curriculum, strategy-based training, practicing only skills that are tested and use of supplementary materials. The studies explained above show that language exams, particularly with the high stakes, profoundly influenced the language teaching mechanisms including the stakeholders (i.e. teachers, students, and policy-makers), classroom dynamics (i.e. content, techniques, and activities), curricular representation (i.e. coverage of all curriculum-related content), language attainment (i.e. achievement of the course objectives), and socio-political issues (i.e. exam marketing, status of languages, and policy-making).

Critical Language Testing Theory

The previous research studies show that language testing exert unique or multi-dimensional effects on the instructional dynamics: such as stakeholders, teaching program and materials, teaching practices. Similarly, use of test results leads to making some important decisions about the students, teachers and programs (Bachman, 1990). Regarding the student-related decisions, Shohamy (1993) highlights that “the placement of students in class levels, the granting of certificates or diplomas”, “the selection of students most suitable for higher-education institutions, and the acceptance of job applicants and program candidates” are the examples of these decisions (p. 5). Such an influential and powerful role of testing is framed as “the power of tests” by Shohamy in 1993 and then theorized as *Critical Language Testing Theory*.

This theory claims that “the act of testing is not neutral. Rather, it is both a product and an agent of cultural, social, political, educational and ideological agendas that shape the lives of individual participants, teachers and learners” (Shohamy, 1998, p. 333). In this way, she brings more comprehensive point of view into language testing beyond psychometric aspects, underlines context-dependent nature of language testing and shows that language testing evolves around political, ideological, cultural and social decisions. Weir (2005) comments on this theory and says that “this view tends to see tests as instruments of power and control” and “their main impact being the imposition of the constraints, the restriction of curricula, providing disciplinary tools for implementing political agenda” (p. 38). It means that testing is critically important in implementing or imposing political decisions on education or curriculum. Shohamy (2017) gives a basic example of critical language testing. When a high-stakes test is designed and introduced to an educational system, stakeholders including principles and students will have demands on test-oriented teaching and practices. Therefore, it will result in a change in the stakeholders’ expectations and aiming at getting higher scores in this test.

To conclude, the life-changing role of language tests is conceptualized by Shohamy as critical language testing. This theory puts emphasis on the political, social, educational and ideological agendas as triggering factors of washback effects. These agendas shape the stakes and purpose of the tests; thereby they influence the stakeholders’ attitudes, perceptions and behaviours.

Language Assessment Literacy and Beliefs

Assessment literacy refers to “knowledge, skills and abilities” that are used in the development, administration and evaluation of any kind of testing or assessment (Fulcher, 2012, p. 125). It also consists of “an individual’s understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions” (Popham, 2011, p. 267). In this respect, teacher assessment literacy interplays with washback effects and then instructional decisions. Stiggins (1991) pays attention to assessment literate individuals and assert that they are familiar with “high-quality assessment”, align assessment methods with goals and objectives, become aware of representing the performance and knowledgeable about unintended results (p. 535).

Along with assessment literacy, teacher beliefs have pivotal importance in shaping assessment practices. For that reason, many research studies are designed to explore language teachers’ assessment beliefs, literacy and practices. Wicking’s research (2017) is on exploring university-level teachers’ assessment beliefs and practices in Japan. According to the research findings, EFL teachers who are not influenced by high-stakes tests are confident in supporting and enhancing their learning through assessment practices. Their positive perceptions lead to effective assessment practices and procedures in the Japanese higher education. Wagl (2011) elaborated on teacher’s role in manifestation of washback effects in the Chinese universities. The study findings displayed that major modifications in test-related and curricular policies might not lead to a change alone in teachers’ methodological practices. These modifications should be reinforced by teacher beliefs, knowledge and perceptions. Hence, in order to increase positive washback effects, EFL teachers should be knowledgeable about test purpose and intended outcomes, and professional training about communicative teaching should be provided, as well. Ha, Tran, and Tran (2021) carried out a research study on language teachers’ beliefs and practices in relation to the curricular movement in Vietnam. The results pointed out that this shift influenced teacher beliefs; thus, these beliefs shaped their assessment practices. They emphasize the significance of in-service training in order to bring about a change in line with curriculum reforms. Concerning pre-service education of English language teaching, Hatipoğlu (2015) investigated expectations and needs of 124 student teachers through questionnaires and surveys. The study findings explain the expectations of improving assessment literacy during English language testing and evaluation training. Additionally, it

is explored that exam-oriented context, testing culture and previous experiences about testing are influential in shaping student teachers' beliefs. Önalın and Gürsoy's (2020) study attempts to explore the use of assessment and results of EFL teachers who work at private and state schools. The results mainly show teachers' perceptions on student involvement, use of alternative assessment and teacher experience. The study emphasized the importance of pre-service and in-service training on language assessment. Büyükkarcı (2016) also calls attention to in-service teacher training on language assessment. By the same token, Taylor (2009) stresses that language teachers should get training when they are included in testing procedures of large-scale tests or "in developing, scoring, interpreting, and improving classroom-based assessments" (p. 24).

Öz and Atay (2017) investigated the assessment literacy of EFL instructors at Turkish universities. They pointed out a mismatch between their perceptions and practices. This study underlines the importance of pre-service education in order for student teachers to practice language assessment. In respect to the assessment practices at state schools in Turkish context, Sarıgöz and Fişne (2018) carried out a research study to develop an understanding about to what extent language assessment practices match curricular requirements in the 4th grades. The basic findings show that there is an emphasis on assessing writing and vocabulary especially with written exams and assignments. However, alternative assessment is less conducted in these grades. Arslan and Üçok-Atasoy (2020) found similar results in lower secondary education. Their study illustrates that EFL teachers follow more traditional methods in assessment than alternative ones although teaching curriculum underlines the use of both formal and alternative assessments.

Educational Policies in Turkey

Table 1 summarizes the basic laws, regulations, and documents that provide a basis for framing teaching and testing policies in Turkey. These official documents are also discussed in the following section more specifically with reference to the classroom-based or large-scale practices of foreign language teaching and testing in each level of education. In 1961, law of education and primary education was put into practice. In this law, the structure and infrastructure of primary education, councils, basic duties, scheduling, student recruitment, continuation and financial issues are illustrated. Basic Law of National Education – 1739, involves aims, principles, formal and informal education, teaching profession, school

structures, instructional materials and duties. In respect to formal education, specific aims that are formulated in line with general aims of National Education are determined for each level of education. Especially, in secondary education, students are guided for higher education and employability in compliance with their abilities and skills as well as societal needs. Also, the cooperation between the MoNE and CoHE is highlighted in order for secondary school graduates to make a transition to higher education. Law of foreign language teaching and Turkish Citizens' learning of different languages and dialects (1983) shape what foreign languages are to be taught and what courses should be given in foreign languages.

In 1997, Turkish National Education witnessed a movement, and eight-year education (Grade 1 to 8) became mandatory and was called "primary education". Upper secondary schools also went through a modification in 2005 in terms of structure and foreign language teaching. With this policy, three-year upper secondary education was transformed into four-year education. Before 2005, there was an implementation of preparatory classes (mainly based on foreign language teaching) at some upper secondary schools, and teaching period in these schools lasted four years. With the official document (No:184) entitled "Restructuring Upper Secondary Education", this implementation was abandoned in those schools, and four-year education was reorganized and rescheduled. In addition, it was determined to adopt foreign language teaching system of Council of Europe (CoE) from primary education to the end of secondary education (see Article 3). In 2006, a regulation specific to foreign language teaching was launched. Official regulation entitled "Ministry of National Education Regulation for Foreign Language Education and Teaching" presents foreign language teaching aims in Article 5. These aims are to enable foreign language learners to have mastery in language skills, communicate in target language and develop positive attitudes to the foreign language. It also makes possible to run supplementary foreign language courses depending on the learners' demands. On the other hand, there is no particular information about how to assess or test language performance. The pre-school and primary education (including lower secondary education) regulation designed more specifically to shed light upon educational practices involves a variety of the topics ranged from schedule and student recruitment to assessment and evaluation of student achievement.

Table 1

*Official Documents and Regulations**

Number	Date	Document Title
10705	12.01.1961	Law of Education and Primary Education
14574	24.06.1973	Basic Law of National Education - 1739
18196	19.10.1983	Law of Foreign Language Teaching and Turkish Citizens' Learning of Different Languages and Dialects
23084	18.08.1997	Change in Law of Education and Primary Education, Basic Law of National Education, Law of Apprenticeship and Vocational Education, Ministry of National Education Organization and Duties
2573/184	07.06.2005	Council of Education and Moral – Restructuring Upper Secondary Education
26184	31.05.2006	Ministry of National Education Regulation for Foreign Language Education and Teaching
27683	03.03.2011	Law of Measurement, Selection and Placement Services
28758	07.09.2013	Ministry of National Education Regulation for Upper Secondary Education
28620	16.04.2013	Regulation for Item Preparation, Item Pooling and Ensuring Item Security
29072	26.07.2014	Ministry of National Education Regulation for Pre-school and Primary Education Institutions
8649147	12.08.2016	Ministry of National Education Regulation for Centralized Examination
6738897	03.04.2019	Ministry of National Education Regulation for Transition to Upper Secondary Education
---	2021	Guide for Application and Administration of Centralized Examination for Upper Secondary Education
---	2021	Guide for Higher Education Examination

*Regulations and laws are listed here in accordance with the publication dates.

To elaborate on the assessment and evaluation, it is explained that assessment should be based on teaching curriculum; non-numerical evaluation should be conducted in the grades 1 to 3; from the 4th grades and onwards, student achievement should be depicted with numerical values: scores. In addition, cut-off score and types of assessment are determined in this regulation. It requires the administration of at least two exams in the grades 4 to 8 and preparation of exam documents and answer keys.

Regarding upper secondary education, the regulation informs the stakeholders about the topics such as scheduling, teaching materials and programs, transition to upper secondary education, assessment and evaluation. It is highlighted that there are three ways of transition to upper secondary schools: without exam, with competency exams or with the scores of the centralized examination. All the procedures and steps are explicated in the guide for transition to upper secondary education. With respect to assessment and evaluation, more specific exam types are listed here: written exams, performance-based exams, skills-based exams and projects. It is expected to develop items in consonance with the curricular objectives. Concerning the foreign language courses, it is emphasized that teaching four skills is essential, and these skills should be assessed through the abovementioned exam types. MoNE devised a regulation for centralized examinations in 2016. This document provides information about the principles of development and administration of the centralized exams that are conducted by MoNE. In the regulation for transition to upper secondary education (MoNE, 2019), test specifications are determined. Accordingly, the curricular objectives specified for the 8th grade will be taken into account in the development of test items. Besides that, diploma score is defined as one of the criteria considered in transition to upper secondary schools. The guide for application and administration of centralized exam for upper secondary education (2021) specifies item types, distribution of items into the subjects, allotted time and coefficients. It is stated that exam will include 90 multiple choice items; 10 items belong to foreign language teaching. Concerning the centralized exam administered at the end of upper secondary education, the guide for higher education institutions (MSPC, 2021) announces that all tests including FLT will include multiple choice items; the number of items for FLT is defined as 80, and allotted time for FLT will be 120 minutes. There is no particular information given about what language objectives and skills are considered and what types of question are developed. When the policies developed within the frame of MSPC, law of measurement, selection and placement

services (2011) maps out MSPC's duties and authorities along with displaying the principles of measurement. In 2013, how to develop items, prepare an item pool and ascertain item security was specified in the regulation based on the previously mentioned law of MSPC. Accordingly, it is elucidated that exams may include open-ended, translation, multiple-choice, listening comprehension and speaking items. Item types are explained in the relevant guides.

The Turkey K-12 Education System: Language Teaching and Testing

Turkish National Educational System comprises two main kinds of education: formal education and non-formal education as explicated in Basic Law of National Education No: 1739. Formal education includes pre-school, primary, secondary (lower and upper secondary) and higher education while non-formal education consists of all educational practices along with or except for formal education and aims to equip the individuals, who continue or leave or never receive formal education, with the instructional, vocational and social skills (MoNE, 2021). In the following sub-sections, the main historical shifts in EFL teaching that fundamentally influence primary and secondary education, duration of language teaching, curriculum-based principles, school-based assessment and large-scale centralized testing are highlighted in detail.

Primary and Lower Secondary Education

English language teaching has gone through several stages in Turkey. Kırkgöz (2007) divides the history of English language teaching into three sequential parts: (1) introduction of English language into Turkish Education System, (2) the major curricular reform in 1997 and (3) the developments in language teaching policies from 2005 onwards. Specifically, the Tanzimat period is the term during which English language was introduced to Turkish education system; the educational reform in 1997 (the implementation of the eight-year compulsory education) led to the development of the new teaching programs; the curricular change in English language teaching in 2006 was carried out to keep up with the globalized education policies and innovations in language teaching. Among these periods, the reform act in 1997 results in drastic changes in language teaching in Turkey. With this educational reform that makes the eight-year education compulsory, English was introduced to primary level (the 4th and 5th grade) as a foreign language course with two lesson hours (Kırkgöz,

Çelik, & Arıkan, 2016). In the meantime, duration of foreign language teaching was four lesson hours in the upper grades (the 6th, 7th and 8th grades) (Genç, 2011). Kirkgöz (2007) asserts that it was the first time that the communicative approach had been presented to English language teaching curriculum. It was also touched upon the learners' attitudes, motivations and strategies and attempted to teach English in the contextualized settings. At the beginning of the 21st century, the field of English language teaching has witnessed the presentation of new approaches and techniques in light of the scientific research findings. Besides, Turkey has continued to develop its relations with European Nations. These developments required the revision of ELT curriculum in use. Accordingly, another curricular revision was conducted by Ersöz et al. in 2006. Through this revision, the pillars of language teaching were theoretically elaborated and their practical implementations are suggested. As for language testing, alternative and authentic way of assessment is emphasized along with the use of European language portfolio (ELP).

While the eight-year education became compulsory in 1997, the 4+4+4 education system was put into practice in 2012, and compulsory education was extended to 12 years covering three levels of education (i.e. primary, lower secondary, and upper secondary education), and the age of English language learning was lowered to 6 and 6,5 years (MoNE, 2013). That is, English language teaching was initiated in the 2nd and 3rd grades with two lesson hours. In consonance with this instructional movement, a new curriculum for English language teaching was devised by the Board of Education in 2013. This curriculum was structured on the basis of the CEFR perspectives and aimed to develop communicative competence. For language assessment, it was intended to promote the use of self-assessment. This curriculum was also revised in 2017 “in line with the pedagogic philosophy of both basic skills and values education”, and this revision includes two dimensions: revising (1) “the theoretical framework in terms of values education, basic skills as themes, and expanding issues such as testing, evaluation, and suggestions” and (2) “target language skills and their linguistic realizations, update of the contexts, tasks and activities, and functions and forms covered” (MoNE, 2018, p. 3). Thus, values education was introduced to English language learners, the framework of language testing and evaluation was shaped by aiming at positive washback, and language-related components such as skills, functions and forms were revised through updating the contexts, tasks and activities.

The current curriculum in use basically follows the CEFR principles. The focus is on the use of language for communicative and authentic purposes. The curricular objectives are to help young learners develop positive attitudes towards English language and motivate them to learn and use English language. With respect to the language skills, speaking and listening are highlighted as the main language skills to be taught at the earlier stages, especially in the 2nd and 3rd grades. In the 4th grades and onwards, it is suggested to enhance the importance of all language skills and components in line with the learners' developmental stages. Regarding duration of language teaching, it is presented that two lesson hours (a lesson hour is 40 minutes) per week are allocated for English language teaching in the primary education (in the 2nd, 3rd, and 4th grades). In lower secondary education, the 5th and 6th graders have weekly three lesson hours of English language. In the 7th and 8th grades, allotted time for language teaching is four lesson hours per week. As for language testing in primary and lower secondary education, it can be stressed that the primary aim of the curriculum is to provide the positive and beneficial washback. Positive washback basically means that language testing supports and improves language learning (Hughes, 2003). In this regard, the current curriculum puts emphasis on the consistency between language learning, teaching and testing. Furthermore, it offers a language testing procedure based on the inclusion of all language skills and components, considering learning styles and learner characteristics, promoting learner autonomy, self-assessment, reflection, and defining learning strengths and weaknesses (MoNE, 2018). In the 2nd and 3rd grades, summative testing techniques are not suggested at all. The primary purpose at these stages is to monitor language learning from a formative perspective. In the 4th grades and onwards, both summative and formative testing techniques are recommended such as alternative assessment, self-assessment, written and oral examinations, homework assignments, and projects.

Regulation for Pre-school and Primary Education Institutions No. 29702 issued by the MoNE in 2014 indicates that the curricular objectives and principles related to assessment and evaluation are taken into account in assessing the learning outcomes. According to this regulation, in the 2nd and 3rd grades, the learner's success is defined with non-numerical values such as “very good”, “good”, or “needs development” in the school reports. On the other hand, numerical values (i.e. points) are used to show the level of success in the 4th grades and onwards. In the 4th grade, the learner success is measured by means of the exams

and in-class participation. In lower secondary education (the 5th to 8th grades), the exams, in-class participation and projects (if available) are taken into account in explicating the learner success. At the end of the 8th grade, the learners, depending on their choice, take a large-scale centralized exam (LGS: Liselere Geçiş Sınavı [Upper Secondary Education Entrance Exam]) that is devised, implemented and reported by the Board of Education in keeping with Regulation for Upper Secondary Education (2013), Regulation for Transition to Upper Secondary Education (2019), and Regulation for Centralized Exam (2016). This exam functions as a medium of selecting the learners and placing them into the upper secondary education. Owing to this function, it has a high-stakes importance for all the stakeholders such as the learners, parents and school administrators. Large-scale exam includes different disciplines with different numbers of items. The centralized exam application and administration guide (2021) points out that there are 10 multiple-choice items for English language discipline, and the coefficient of English language test is 1 whereas Turkish, Mathematics and Science disciplines' coefficient is 4. This 10-item English language test is a part of the social sciences test with 50 items, and the total time for this social science test is 80 minutes. From a wider perspective, English language test, which is a part of the large- scale centralized examination, plays a pivotal role in shaping educational decisions including transition to upper secondary education and has important consequences for the learners' academic development. The 8th graders, who do not prefer taking this large-scale exam, can apply for upper secondary schools who select the students on the basis of the students' home address and diploma scores.

Upper Secondary Education

In Turkey, upper secondary education covers the 9th, 10th, 11th and 12th grades. It is compulsory for Turkish citizens to attend and complete this level of education (see Basic Law of National Education Article 26). When the instructional settings of upper secondary education were examined, it is seen that there are different types of schools that are established according to the employability purposes to be achieved at the end of educational period. In the Regulation for Upper Secondary Education, twelve types of schools are identified within the framework of the institutions with general-purposes and vocational-purposes as follows: (1) Science, (2) Social Sciences, (3) Anatolian, (4) Fine Arts, (5) Sport, (6) Anatolian Religious, (7) Vocational and Technical Anatolian, (8) Multi-program

Anatolian, (9) Special Education Vocational Upper Secondary Schools, (10) Vocational Education Centers, (11) Vocational and Technical Education Centers and (12) Education Campuses

After the 8th graders have graduated from lower secondary schools (i.e. lower secondary schools or religious lower secondary schools), they are offered two options as explained in the previous section: (1) entering LGS (centralized exam) and being placed into the upper secondary schools that accept the students with exam scores or (2) not taking LGS (centralized exam) and registering for upper secondary schools that accept the students on the basis of the home address and diploma scores. It means that upper secondary schools are also divided into two categories according to selection of the students. The list of the schools that can be placed into only via centralized exam scores is provided in the Guide for Application and Administration of Centralized Exam (LGS) each year.

Similar to primary and lower secondary education, EFL teaching in upper secondary education has been widely influenced by the historical shifts taken place in education. During 1953-1959 years, upper secondary colleges were established in different parts of Turkey in order to teach the main courses like mathematics and sciences in the medium of English language (Güven, 2018). Then in 1975, the school name of these colleges were transformed into “Anatolian Upper Secondary Schools” and the number of these schools increased in the 1990s (Çakır, 2017). There were language preparatory classes (intensive courses of English) in Anatolian upper secondary schools and some other types of schools in which twenty-four lesson hours were devoted for teaching English language (Yücel, Dimici, Yıldız, & Bümen, 2017). However, in 2005, the implementation of preparatory education was abolished by Board of Education and Discipline in line with the new regulation entitled “Restructuring Upper Secondary Education” (Document No: 184; Date: 07.06.2005) and the hours of English language lessons were reorganized in accordance with the level of teaching. Weekly duration of English language teaching was scheduled as ten lesson hours in the 9th grades whereas four lesson hours were allocated for the 10th, 11th and 12th grades (Kırkgöz, 2007). Then in 2012, the 4+4+4 system was introduced to the Turkish educational system, and this structural change and newly designed ELT curriculum for primary and lower secondary education required the ELT curriculum to be revised in upper secondary education, too. So, the revisions were conducted respectively in accordance with the latest developments in English language teaching and innovations in the Turkish education system.

The current ELT curriculum, based on the principles of the CEFR, embarks on the authentic use of language in a communicative way by using technology, promoting learner autonomy and collaboration, and increasing reflective thinking with four lessons hours allocated for English language teaching in each grade (MoNE, 2018). Drawing attention to the expansion of the technology use in the adolescents' life, the curriculum attaches paramount importance to the integration of technology with language teaching and assessment. It also emphasizes the embeddedness of values education into language themes and units. Another curricular focus is on the operational use of communicative competence in the language classes. It is stated that the washback effect of language testing restricts the development of the communicative competence except for the rise of grammatical competence. In order to enhance communicative competence and eliminate the mismatch between teaching and testing, authentic assessment is highly recommended in upper secondary education. In this level of education, assessment is conducted by means of "the mixture of alternative, traditional, and electronic assessment types" (MoNE, 2018, p. 14). Discussion time activities, video blogs, pen-and-paper exams, e-portfolios, and teacher observation are the examples of assessment practices. Feedback is another critical component of the assessment and evaluation. Teacher, parent, peer, self- and computer mediated feedback are the main sources of feedback that learners can get about their language learning (p. 15). Regulation for Upper Secondary Education (MoNE, 2013) frames how to conduct assessment and evaluation at upper secondary schools. Accordingly, written exams, performance-based assessment and projects that are prepared in line with the curricular perspectives are the main sources to gather information about the learners' achievement. For foreign language lessons, assessing four skills are explicitly emphasized. Formative aspects of the assessment are highlighted, and higher-order thinking skills (e.g. critical thinking and problem-solving) are taken into consideration in the assessment and evaluation. It is expected to ensure reliability, validity and practicality of the assessment tools. Regardless of how many lesson hours are allocated, it is required to administer two written exams at least in each discipline. In addition to written exams, performance-based assessment, projects and overall judgment (i.e. participation, attitude) are taken into account in defining the learners' level of attainment in the target area.

At the end of the 12th grade, the learners aiming at getting university-level education take a large-scale centralized exam called the Exam for Higher Education Institutions (YKS). It has

five different components, one of which is the Foreign Language Test (YDT). Firstly, all the examinees take the Basic Competency Test (TYT) including 120 items in the disciplines of Turkish language, mathematics, science and social sciences.² Then they take the second exam which is related to their field of study. The 12th graders, who want to study English language at the tertiary level, take the Basic Competency Test and Foreign Language Test in different sessions. Foreign language test includes 80 multiple choice items and basically focuses on vocabulary, grammar and reading comprehension.



² After completing the 10th grade, the learners make a decision on their field of study which is influential in the coverage of content and determines the type of large-scale exam to be taken.



CHAPTER III

METHODOLOGY

This chapter covers the methodological procedures used in the current research study. First of all, the research aims and questions are presented to guide data collection and analysis. Then the research design and sampling method are theoretically and practically illustrated. Next, data collection processes and instruments are explained step by step. Finally, how qualitative data were analyzed is discussed with reference to the analysis techniques.

Research Aims and Questions

This study aims to investigate the washback effects of English language exams on the dynamics of EFL teaching and learning at the state schools (primary, lower secondary and upper secondary schools) from language teachers' perspectives. Firstly, there is an attempt to elaborate on the effects of teacher-based assessment and large-scale centralized testing on the instructional settings in line with EFL teachers' teaching and testing practices and beliefs. Then it is planned to explore the triggering forces of these washback effects at macro- and micro-level.

In order to achieve these aims, the study focuses on the following research questions:

1. How does language testing influence EFL teaching dynamics at the state schools?
 - 1.1. What are the washback effects of low-stakes language assessment in primary and secondary education?
 - 1.2. What are the washback effects of high-stakes language testing in primary and secondary education?

2. What triggers the washback effects of language testing on EFL teaching dynamics at state schools?

2.1. What are the triggers that exert the washback effects of low-stakes language assessment in primary and secondary education?

2.2. What are the triggers that exert the washback effects of high-stakes language testing in primary and secondary education?

These research aims and questions pave the way for framing research design, designing data collection instruments and reporting data analysis results. As illustrated in the chapter of *Introduction*, EFL teaching dynamics here refers to teaching and testing content, methodology, materials, and significant others like students, parents, teachers, and administrators. In this regard, EFL teachers' views, teaching and testing practices and beliefs as well as Field Experts' interpretations of the findings will be used as benchmarks to unveil test effects, mediating factors and triggering forces.

Research Design

This study adopts the multiple case study as a research design within the framework of the qualitative research methodology. Qualitative research deals with the holistic description of the phenomenon by giving all the details in a particular setting (Fraenkel, Wallen, & Hyun, 2012). Sherman and Webb (1988) put forward that “the aim of qualitative research is not verification of a predetermined idea, but discovery that leads to new insights” (p. 5). In this sense, qualitative research enables the researchers to discover the nature of the construct under the study by investigating the relationships between the situations or portraying the interaction among the variables. There are five fundamental features of qualitative studies (Bogdan & Biklen, 2007).

1. It is based on natural inquiry. The research is conducted on the natural setting of the phenomenon, and “the researcher is the key instrument” (p. 4).
2. It deals with descriptive data. As the aim of the research is to describe a particular setting, the main focus is on collecting, analyzing and interpreting the descriptive data.

3. It is process-oriented. Rather than emphasizing only product, it is “concerned with process” (p. 6)
4. It approaches the data with an inductive method. The process starts with investigating the small units of the phenomenon and then a holistic/general description is obtained.
5. The focus of attention is on meaning. It basically intends to make sense of what, how and why the phenomenon occurs.

To summarize, qualitative studies examine the natural setting inductively by providing rich descriptive data about the phenomenon, presenting the meaning of the relationships in the given situations, and depicting the process of the events/activities. In line with these pillars of qualitative methodology, this study makes sense of the descriptive data inductively and sharpens an understanding about how the washback effects are exerted in primary and secondary education.

Case study is a research design that is paradigmatically shaped by the qualitative research methodology. It describes an event, activity or situation in its natural environment. From the terminological standpoint, case study is “a method of collecting and organizing data so as to maximize our understanding of the unitary character of social being or object studied” (Dörnyei, 2007, p. 152). A case could be an individual, group, program, institution, community or phenomenon (Gillham, 2000; McKay, 2006; Yin, 2003). Bloor and Wood (2006) explain the aim of the case studies as “to understand social phenomena within a single or small number of naturally occurring settings” (p. 27). Stake (2006) states that the researchers are required to “generate a picture of the case and then a portrayal of the case for others to see” in the case studies (p. 8). That is to say, they not only broaden the understanding of the nature of the phenomenon but also give a holistic picture of the case by indicating the links and interactions between the units.

There are some strengths of using the case study as a research design. It enables the researchers to concentrate on a specific case, which may be difficult in a large-scale study (Mackey & Gass, 2005). When the pivotal emphasis is placed on a specific case, all the underlying factors and components of the case can be characterized in detail. It also helps the researchers get “a thick description of a complex social issue embedded within a cultural context” (Dörnyei, 2007, p. 155). Thus, the nature of complex issues can be richly depicted. On the other hand, there has been a concern about the application of the case studies in terms

of the generalizability of the results. At this point, Yin (2003) draws attention to the distinction between different research paradigms regarding the concept of generalization. Accordingly, the case study attempts to make analytical generalization, which refers to providing research findings that can be “generalizable to theoretical propositions” (p. 10). So, transferability of the results gains importance in the qualitative studies (Yıldırım & Şimşek, 2013).

Yin (2003) explains five components of the case study research (p. 21): *Questions* (‘How’ and ‘why’ questions are addressed in the case studies), *Propositions* (“to direct attention to something” to be examined), *Unit of analysis* (“fundamental problem of defining what the case is”), *The logic linking the data to the propositions* and *The criteria for interpreting the findings*. When the components of the current study are examined, it can be stated that it aims to answer the question of *how washback effects occur in macro and micro dynamics of EFL teaching in Turkey* and *why they influence these dynamics*. In this respect, teacher-based assessment and large-scale testing are investigated to understand the nature of these effects. The cases and units of analysis are given in Figure 1:

Primary Education	Lower Secondary Education	Upper Secondary Education
•Teacher-based Assessment •Centralized large-scale testing	•Teacher-based Assessment •Centralized large-scale testing	•Teacher-based Assessment •Centralized large-scale testing

Figure 1. Cases and units of analysis in the current research study

There are four different case study designs: (1) single-case (holistic) designs, (2) single-case (embedded) designs, (3) multiple-case (holistic) designs, (4) multiple case (embedded) design (Yin, 2003). In multiple-case embedded designs, there are multiple cases and units of analysis. As can be seen in Figure 1, this research adopts multiple-case embedded design since there are three different cases (i.e. primary, lower secondary, and upper secondary education) and two units of analysis (i.e. teacher-based assessment and large-scale testing) in each case. Stake (2006) asserts that “an important reason for doing the multicase study is to examine how the program or phenomenon performs in different environments” (p. 23).

Accordingly, this study aims to analyze how testing mechanisms and their effects operate in different levels of education.

Sampling

Sample refers to the people or phenomena that represent the universe (Yıldırım & Şimşek, 2013). The process of selecting these people and phenomena from the universe is called sampling. This study uses two main methods of purposeful sampling: maximum variation sampling and criterion sampling. “The logic and power of purposeful sampling lie in selecting information-rich cases for study in depth” (Patton, 2002, p. 230). The participants of this research are EFL teachers who work at primary and secondary schools in the MoNE and field experts who are EFL teacher trainers and have expertise in the areas of language testing and language teaching policies. Maximum variation sampling was employed to select EFL teachers from different regional and educational settings. In this sampling, the purpose is to reflect maximum variability of the individuals (Yıldırım & Şimşek, 2013). Hence, this study collected data from 25 EFL teachers working in different cities in Turkey. These teachers have different contextual variables in foreign language instruction and various backgrounds of the pre-service education in which they gained experiences in the field of language testing. The main purpose to collect data from the EFL teachers about the washback effects is that they are the pivotal agents of language teaching at schools and are primarily affected by testing-related factors. At the same time, they influence instructional dynamics as a result of testing-based decisions.

As for the criterion sampling, it simply means choosing the participants who basically meet a criterion set within the scope of the study (Dörnyei, 2007). 9 Field Experts/EFL teacher trainers were selected with this kind of sampling. The key criterion here is to hold academic studies in the fields of English language testing or language teaching policies. The logic behind sampling EFL teacher trainers to elaborate on the triggers related to the washback effects is that those scholars have in-depth knowledge, expertise, experiences, practices and insights about language testing and teaching. In this way, it is intended to shed light upon underlying factors that govern the impacts of language testing by interviewing with EFL teacher trainers.

Participants

In the current study, there are two separate subject groups taking active roles in English language teaching and testing in Turkey. The first group is English language teachers who work at different levels of education in MoNE. The second group is Field Experts who work at tertiary level and train pre-service teachers to get them qualified to teach and assess English language in various educational contexts. The demographics of the participants are given below.

EFL Teachers

In this research, preliminary interviews were conducted with 25 EFL teachers working in MoNE. The characteristics of EFL teachers are presented in Table 2. These teachers work in different levels of education. That is to say, 4 teachers work in primary education (PST), 5 teachers work in lower secondary education (LST) (two of them work at religious lower secondary schools), 3 teachers work in both primary and lower secondary education (PLST), and finally 13 teachers work in upper secondary education (UST) (nine of them work at Anatolian upper secondary schools, two of them work at vocational upper secondary schools, and two of them work at science upper secondary schools). The reason why the number of the participants is relatively higher in upper secondary education is that there are twelve different types of school at this level (MoNE, 2013), and the aim is to ensure maximum variability.

There are three main BA (Bachelor's Art) programs from which EFL teachers graduated. A majority of teachers (n=22) graduated from English language teaching (ELT) department. While two of them have a BA degree at English language literature, one teacher graduated from English linguistics department. There is a diversity of the universities in which EFL teachers completed the BA degree. These institutions are Gazi University (n=7), Yıldız Technical University (n=5), Atatürk University (n=2), Erciyes University (n=2), Middle East Technical University (n=1), Selçuk University (n=1), Ondokuz Mayıs University (n=1), Anadolu University (n=1), Ankara University (n=1), Hacettepe University (n=1), Uludağ University (n=1), İstanbul University (n=1), and Mugla Sıtkı Kocaman University (n=1).

Considering the gender of the participants, 7 teachers are male teachers, and 18 teachers are female. They are aged from 25 years to 46 years. The mean of age is 31 years. As for the

participants' teaching experiences, it can be outlined that EFL teachers are experienced in language teaching between 2 and 24 years. The mean of teaching experience is 7 years at primary schools, 8 years at lower secondary schools, 7 years at both primary and lower secondary schools and 10 years at upper secondary schools. The diversity of the pre-service education is another criterion in choosing the participants because it is aimed to understand testing variables that are originated in different educational backgrounds. With respect to the MA degrees, 7 teachers graduated from various MA programs (4 teachers =ELT, 1=Instructional Technologies, 1=English Language Literature, 1=Translation and Interpreting). On the other hand, 5 teachers are currently pursuing their MA degree in educational and language-related programs (2=ELT, 2=Educational Sciences, and 1=Teaching Turkish as a Foreign Language)

Marmara Region	• İstanbul, Kocaeli, Bursa
Black Sea Region	• Bartın, Samsun, Tokat, Artvin
Eastern Anatolia Region	• Erzurum, Van, Bitlis
Aegean Region	• Muğla, Denizli, Manisa
Central Anatolia Region	• Ankara, Konya, Yozgat, Sivas, Kayseri
Southeastern Anatolia Region	• Adıyaman, Gaziantep, Mardin
Mediterranean Region	• Burdur, Adana

Figure 2. Regions and cities in which EFL teachers work

Figure 2 shows the regions and cities in which EFL teachers work. It was attempted to get sampling from seven regions of Turkey and 23 different cities located in those regions. The reason why the participants were selected from seven different regions is that washback is a complex phenomenon (Watanable, 2004) and displays different outcomes depending on the educational settings. In order to capture the whole picture of the operationalization of washback in the Turkish MoNE, various contexts situated in the regions of Turkey were included in the analysis.

Table 2

Characteristics of the Participants (EFL Teachers) continued

No	Participant	School Type	Region	Grades*	Teaching Experience	Degree & Field	Age	Gender
1	PST-1	Primary	Southeastern Anatolia Region	2-3-4	7 years	BA – English Language Teaching MA – English Language Teaching	29	Male
2	PST-2	Primary	Eastern Anatolia Region	2-3-4	5 years	BA – English Language Teaching	29	Male
3	PST-3	Primary	Central Anatolia Region	2-3-4	2 years	BA – English Language Teaching MA – Educational Sciences (going on)	36	Female
4	PST-4	Primary	Central Anatolia Region	2-3-4	13 years	BA – English Language Teaching BA – Sociology (going on)	35	Female
5	PLST-1	Primary & Lower Secondary	Marmara Region	2-3-4-5-6-7	1.5 years	BA – English Language Teaching	25	Female
6	PLST-2	Primary & Lower Secondary	Aegean Region	2-3-4-5-6-	8 years	BA – English Language Teaching	31	Female
7	PLST-3	Primary & Lower Secondary	Black Sea Region	1-2-3-4-5-6-7-8	10 years	BA – English Language Teaching MA- English Language Teaching (not completed)	32	Female
8	LST-1	Lower Secondary	Marmara Region	5-8	8 years	BA – English Language Teaching	30	Female
9	LST-2	Lower Secondary	Black Sea Region	5-6-7-8	7 years	BA – English Language Teaching	30	Female
10	LST-3	Lower Secondary	Central Anatolia Region	6	10 years	BA – English Linguistics	35	Male
11	LST-V-1	Lower Secondary	Marmara Region	5-6-7-8	8 years	BA – English Language and Literature	30	Female
12	LST-V-2	Lower Secondary	Marmara Region	5-6-7-8	8 years	BA – English Language Teaching MA – Teaching Turkish as a Foreign Language (going on)	30	Female
13	UST-A-E-1	Upper Secondary	Mediterranean Region	9-10	15 years	BA – English Language Teaching MA - English Language Teaching (going on)	37	Female

No	Participant	School Type	Region	Grades*	Teaching Experience	Degree & Field	Age	Gender
14	UST-A-E-2	Upper Secondary	Central Anatolia Region	9-10-11-12	5 years	BA – English Language and Literature	33	Female
15	UST-A-E-3	Upper Secondary	Eastern Anatolia Region	9	19 years	BA – English Language Teaching MA – English Language and Literature	41	Female
16	UST-A-WE-1	Upper Secondary	Aegean Region	9-10-11-12	13 years	BA – English Language Teaching	36	Female
17	UST-A-WE-2	Upper Secondary	Southeastern Anatolia Region	9-10-11-12	2 years	BA – English Language Teaching MA - English Language Teaching (going on)	27	Male
18	UST-A-WE-3	Upper Secondary	Southeastern Anatolia Region	9-10-11-12	8 years	BA – English Language Teaching MA – Information Technologies and Instruction	30	Male
19	UST-A-WE-4	Upper Secondary	Central Anatolia Region	9-10-11-12	8 years	BA – English Language Teaching	30	Female
20	UST-A-WE-5	Upper Secondary	Aegean Region	9-10-11-12	24 years	BA – English Language Teaching MA – Translation & Interpretation	46	Male
21	UST-A-WE-6	Upper Secondary	Mediterranean Region	11	7 years	BA – English Language Teaching MA – Educational Sciences (going on)	33	Female
22	UST-S-E-1	Upper Secondary	Black Sea Region	11-12	17 years	BA – English Language Teaching MA – English Language Teaching	35	Female
23	UST-S-E-2	Upper Secondary	Eastern Anatolia Region	11-12	5 years	BA – English Language Teaching	30	Female
24	UST-V-E-1	Upper Secondary	Black Sea Region	9-10-11-12	6 years	BA – English Language Teaching MA – English Language Teaching	28	Female
25	UST-V-WE-2	Upper Secondary	Central Anatolia Region	9-10-11	5 years	BA – English Language Teaching MA – English Language Teaching	27	Female

*Grades indicate the levels in which EFL teachers taught English during the term the study was being conducted. EFL teachers may have experiences in other grades from the previous years.

Field Experts

Field experts are the scholars who work as EFL teacher trainers at the different universities in Turkey. Their academic publications are on English language testing and language teaching policies or make contributions to the development of these fields. From instructional perspectives, they train pre-service language teachers about how to teach and test English language in different levels of education. As regards the demographics of the field experts, 7 of them are male, and 2 experts are female. Their age is between 41 and 68 years. The mean age is 50. Their teaching experiences are ranged from 16 to 41 years. 4 experts worked in MoNE as an EFL teacher. 8 experts taught English language testing at the BA-level courses, 7 experts gave various seminars to in-service language teachers about language testing. Figure 3 illustrates the private (n=1) and state (n=8) universities at which the field experts work.

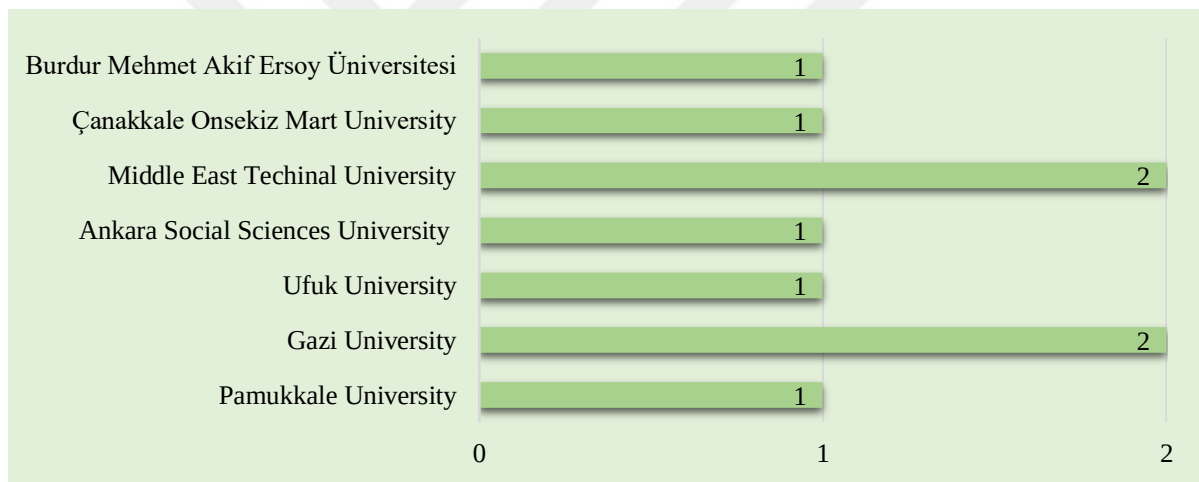


Figure 3. Distribution of the universities at which the field experts work

As pointed out in Figure 3, field experts work at different universities located in the cities of Ankara, Burdur, Çanakkale and Denizli. When the experts' educational backgrounds are examined, it is seen that all the experts completed their BA degree in the ELT programs. 2 experts have additional BA degrees in the department of linguistics. As for the MA level, all the field experts hold their MA levels in ELT, and one expert has the additional MA degree in linguistics. In regard to the PhD levels, 7 of them have their degrees in ELT, and two of them completed their PhD in different fields: Linguistics and Curriculum & Instruction.

Data Collection

Case study design involves a variety of the data collection sources ranging from observation to interviews. This study mainly utilizes the method of interviewing to get in-depth data about the research focus. Interviewing offers the researchers to obtain rich information about the research constructs from the respondents' points of view. Basically, interview refers to "the elicitation of research data through the questioning of respondents" (Bloor & Wood, 2006, p. 104). Through interviews, the researchers can take part in the process in which the individuals' behaviours are shaped (Seidman, 2006) and develop a comprehensive understanding of the individuals' interpretation (Bogdan & Biklen, 2007). Mackey and Gass (2005) accentuate some advantages of conducting interviews. They help the researchers collect data about a research variable that cannot be directly observed. Furthermore, they provide opportunities for the researchers to elaborate on the research constructs that are not explained enough or require more clarity.

There are different types of interviews used in the educational research. Dörnyei (2007) classifies the interviews into three types according to their degree of structure: structured interviews, semi-structured interviews and unstructured interviews. This study gathers qualitative data by means of semi-structured interviews. In semi-structured interviews, there is "a written list of questions" that is used as a guide to maintain asking and responding procedure, and it helps the researchers deepen the topic that needs further information (Mackey & Gass, 2005, p. 173). It is appropriate to use this kind of interviewing in the situations when the target "phenomenon" or "domain" is adequately understood, and there is a sufficient number of the questions developed, but it is not intended to use "ready-made response categories that would limit the depth and breadth of the respondent's story" (Dörnyei, 2007, p. 136). Gillham (2000) states that "the semi-structured interview is both flexible, and at the same time, standardized" (p. 69). Put another way, in semi-structured interviews, there are standard written questions which both enable the researchers to sharpen their understanding of the topic and give flexibility to them in elaborating on the responses requiring further insights.

In this research study, two different semi-structured interviews were conducted with two different samples sequentially: EFL teachers and field experts. Interview forms are given in Appendix A. How the questions in the semi-structured interview forms are developed, and

how these forms are validated are explained below in the following section of Data Collection Tools:

Data Collection Tools

Interview Form for EFL Teachers

This form has two main parts and an introductory part. In the introductory part, the researcher introduces herself, gives information about the research aims and explicates the research constructs. She puts emphasis on the confidentiality of the obtained data and expresses the estimated time of interviewing. In Part 1, demographic questions are directed to the interviewees. These questions are about their age, gender, program of graduation, teaching experience, grades they teach English, and testing backgrounds. In Part 2, interview questions are listed in accordance with the content and focus of this research study. Table 3 summarizes the alignment of the interview questions with the research questions and variables. As can be seen in Table 2, the interview form has 15 questions given under two main categories: teacher-based assessment and large-scale centralized language testing. Research variables are English language teaching methodology, assessment and evaluation, and the stakeholders. In each question, alternative questions and probes are provided in relation to the theoretical bases of the constructs. Patton (2002) elucidates that “probes are used to deepen the response to a question, increase the richness and depth of responses, and give cues to the interviewee about the level of response that is desired” (p. 372). In this regard, the probes enable the researcher to expand the points that need further elaboration. There are six types of questions that are used in the interviews (Patton, 2002). These are experience and behaviour questions, opinion and values questions, feeling questions, knowledge questions, background and demographic questions, and sensory questions. In this semi-structured interview, background/demographic questions, experience questions and opinion questions are addressed to EFL teachers. Patton (2002) also asserts that good interview questions should be open-ended, neutral, singular and clear. These criteria are taken into consideration in the development of the interview questions. More importance is given to the inclusion of open-ended, one-dimensional, understandable and clear questions in the interview forms.

Table 3

Linking Interview Questions for EFL Teachers to the Research Variables

Unit of Analysis	Research Question	Research Variables	Interview questions
Teacher-based Assessment	R.1.1	Language teaching methodology (language objectives, content, teaching methods, practices, materials)	Q2-Q3-Q4
	R.1.1 R.2.1	Assessment methods and practices, use of the assessment results	Q1-Q8
	R.1.1 R.2.1	Stakeholders (teachers, students, parents, school administrators, colleagues)	Q5-Q6-Q7
	R.1.2 R.2.2	Language teaching methodology (language objectives, content, teaching methods, practices, materials)	Q9-Q10
Large-scale centralized testing	R.1.2	Assessment methods and practices, use of the assessment results	Q11-Q15
	R.1.2 R.2.2	Stakeholders (teachers, students, parents, school administrators, colleagues)	Q12-Q13-Q14

In the development process of the semi-structured interview form, some sequential steps are closely followed as given in Figure 4: a) review of literature, b) the first draft of the interview form, c) piloting, d) expert opinions, e) revision, f) preparation of the finalized form. Firstly, the relevant literature was elaborately reviewed, and the main research variables and constructs were specified. In line with this review and the research questions, the first draft of the interview form was prepared. The questions mainly investigated two constructs: teacher-based assessment and large-scale centralized testing. Based on these two constructs, sub-questions that focus on the impact of the assessment on teaching, learning and stakeholders such as teachers, students, parents and school administrators were written. After the preparation of the first draft, the questions were piloted in a small group of the participants (N=6). Two EFL teachers in each level of education (primary, lower secondary and upper secondary, in total 6 teachers) were interviewed through the phone conversation,

and interviews were recorded with note-taking. Piloting results showed that more introductory questions which elicit EFL teachers' assessment practices and knowledge should be placed into the interview form. The variable of the stakeholders should address more participants taking part in the English language testing. Not only the effects of the examination on the stakeholders but also their expectations from the assessment/testing procedure should be questioned. EFL teachers' in-class teaching and testing practices should be exemplified with the prompts. Each grade in every level of education should be examined in detail in terms of the nexus between curricular principles and in-class implementation, and its relation to language testing. Teacher's role in society and education should be discussed in light of the language testing policies.

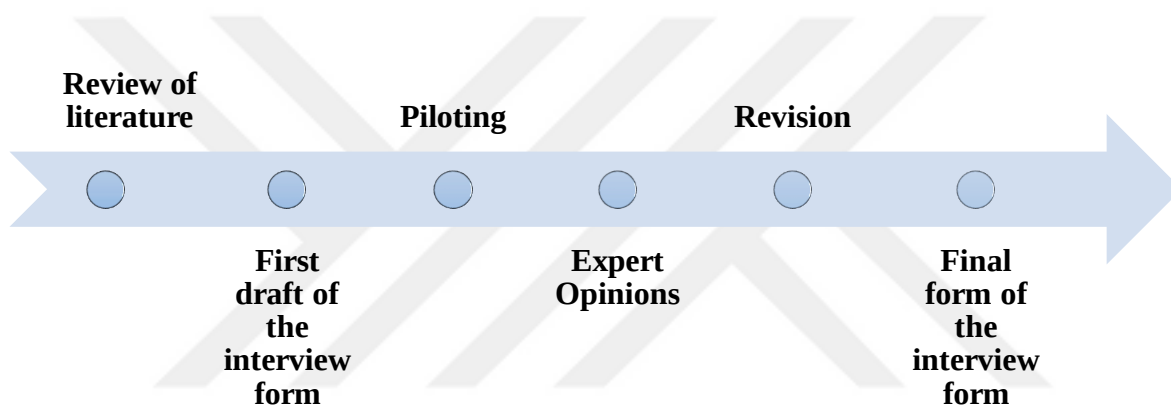


Figure 4. The development process of the semi-structured interview form

The interview questions that are related to the issues explained above were developed and put into the semi-structured interview form. In the fourth step, this revised form was sent to 10 experts (5 experts in the department of educational measurement and 5 experts in the department of English language teaching) in order to evaluate the content and construct validity of the interview questions. 10 experts examined each interview question and stated their opinions about the appropriateness of the questions to the content of the study (see Appendix B). After getting the experts' feedback, a major revision was carried out in the interview form. First of all, the variable of language teaching methodology was clarified and expanded. New alternative questions and probes were developed and placed under each main question. Wording of the questions was improved in terms of the lexical diversity and grammatical accuracy. In the sixth step, demographic questions and an introductory part were added to the interview form, and the understandability of the interview questions were

examined by an EFL teacher. With the help of this final feedback, some questions were reorganized, probes were enriched, and interview form was finalized (Appendix A).

Interview Form for Field Experts

After interviewing with EFL teachers, semi-structured interviews were conducted with the field experts in the following step. EFL teachers' interview responses guided the development of the interview questions to be conducted with the field experts. In other words, teacher interviews conceptualized how washback effects function in primary and secondary education and clarified the underlying factors triggering washback effects. Those factors were discussed with the field experts within social, political and economic frameworks.

Table 4

Linking Interview Questions for Field Experts to the Research Variables

Interview questions	Preliminary Findings	Research Questions	Research Variables
Q1	Mismatch between curricular perspectives and centralized language testing	R.2.1 R.2.2	Testing four language skills Language testing policy Discrepancy between testing and curriculum
Q2	Regulation-governed in-class teaching and testing practices	R.2.1 R.2.2	Language teaching and testing regulations in primary and secondary education Regulation- and policy-making
Q3	Teacher-related factors in teaching and testing	R.2.1	Diversity in teacher behaviours Teachers' testing and teaching beliefs, knowledge, missions
Q4	Stakeholders' perceptions on language testing and teaching	R.2.1 R.2.2	Social perspectives about language teaching and testing, societal needs for testing
Q5	Power of language tests in society	R.2.2	Stakeholders' motivation, achievement, social and economic reasons and consequences

The preliminary findings of EFL teachers' interviews were related to the grade-based shifts in English language teaching and testing in each level of education, discrepancy between testing and teaching, teacher-related factors (e.g. missions, beliefs), stakeholders' perceptions and instructional factors (e.g. regulations and allocated time). Those findings pathed way for the development of EFL teacher trainers' interview questions as outlined in Table 4. As highlighted in Table 3, semi-structured interview form for field experts has five main questions, subordinate alternative questions and probes. The form includes an introductory part, which underlines the research aim and focus, and demographic questions (Appendix A). Demographic, experience and opinion questions are the main types of questions included in the form. In addition, clarity, neutrality, open-ended inquiry and singularity were taken into account in the development of the interview questions (Patton, 2002). Those questions were directed to 9 field experts, and qualitative data were analyzed as depicted in the following section.

Data Analysis

The qualitative data collected within the multiple case study research were analyzed with the method of content analysis. It is the method used "for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use" (Krippendorp, 2004, p. 18). Through content analysis, a wide range of data sources from textbooks, academic papers to advertisements and public speeches can be examined (Fraenkel, Wallen, & Hyun, 2012). Considering all the communication sources, content analysis is carried out in line with some certain procedures as elucidated below (Cohen, Manion, & Morrison, 2007; Krippendorp, 2004):

- (a) Specifying the research questions and constructs,
- (b) Identifying the data sample,
- (c) Defining the context and analysis unit,
- (d) Coding the qualitative data,
- (e) Categorizing, classifying and comparing the codes and data,
- (f) Making inferences and drawing conclusions

Especially in the analysis of the recorded spoken data, Dörnyei (2007) proposes four basic steps: (1) transcribing the data (2) pre-coding and coding, (3) growing ideas and data display, and (4) interpretation and drawing conclusions (p. 246). Likewise, Yıldırım and Şimşek (2013) elaborate on the qualitative data analysis procedures and pose four steps in this process: (1) coding the data, (2) specifying themes, (3) revising codes and themes, and (4) interpretation. In keeping with the steps given above, the qualitative data of this research were transcribed immediately after the data collection; they were coded in two phases (pre-coding and second-level coding); categories and themes were specified and then revised; the raw data and data-driven findings based on the relationships between the codes, categories and themes were displayed; and the data were interpreted in consideration of the research questions and constructs.

To begin with, the qualitative data gathered from semi-structured interviews conducted with 25 EFL teachers and 9 Field Experts were recorded with the consent of the participants. Table 5 points out how much time these interviews last in each case. As explicated there, 28.2 hours (1692 minutes) were allocated to interview with the teachers and experts. Then those audio records were systematically and verbatim transcribed. Almost each interview was entirely transcribed in order to include all the details that may give insights to research constructs and provide a basis for building relationships between these constructs.

Table 5. *Interviewing Duration*

Level of Education	Primary Education Teachers	Lower Secondary Education Teachers			Upper Secondary Education Teachers			Field Experts	Total
LE	PT	PLST	LST		USE-A	USE-S	USE-V	FE	T
Time (minute)	150m	103m	403m		526m	49m	58m	510m	1692m

The data not irrelevant to the research scope were not included in the transcription and therefore data analysis. Miles and Huberman (1994) underscore the importance of three aspects in qualitative data analysis, and one of them is data reduction that basically refers to “the process of selecting, focusing, simplifying, abstracting and transforming the data that appear in written-up field notes or transcriptions” (p. 10). During the transcription and initial coding process, data reductions were made in line with the research constructs; thus, the qualitative data were prepared for coding procedures.

As the essential step of the content analysis, coding refers to “denoting concepts to stand for meaning” (Corbin & Strauss, 2015, p. 76). Yıldırım and Şimşek (2013) explain that coding can be conducted through assigning a keyword, phrase or metaphors to the concepts. During this stage, the coders analyze the data by reflecting upon their own reasoning about the variables (Miles & Huberman, 2004). Accordingly, the transcribed data were thoroughly examined in order to find out the key points emerged from the text and related to the units of analysis. Two basic kinds of coding were employed in the current study: initial coding and second-level coding. In initial coding, transcripts were read thoroughly and coded manually. Then the documents in each case were respectively imported into the software program of NVivo 12 Plus that provides the researchers with opportunities to manage, query, visualize and report the qualitative data (Bazeley & Jackson, 2013). In this program, manual codes were transferred into the nodes and checked again through this process. During initial coding, the number of nodes generated in NVivo is specified as 207.

After initial coding was finalized, all the codes were revised in terms of their similarities and differences. Following these revisions, the second-level coding process was initiated, and the number of codes was reduced to 106 in regard to the code interdependency and consistency. At the end of the coding process, 106 codes were grouped into 14 categories in three cases as seen in Table 5 (PT=24, LST=38, UST=44). These categories are not only developed by the research constructs but also emerged from the content of the codes. Patton (2002) pays attention to two main criteria in defining the categories: “internal homogeneity” and “external heterogeneity”. The first criterion means that the data given in a category should be internally consistent with the other data in the same category. As for the second criterion, it is underlined that there should be clear and obvious discrepancies between the main categories in the same analysis. During the data analysis, it was attempted to ensure these two criteria by inserting the codes with the similar meanings into the same category (internal homogeneity) and comparing the newly formulated categories with others to avoid multi-connection problems. For example, within the case of primary education, the codes of *teacher beliefs*, *teacher identity* and *teacher missions* were grouped into the category of teacher-related factors. On the other hand, the codes of *administrators*, *colleagues*, *parents and students’ expectations* were categorized as the stakeholders’ expectations. The differences between these two categories are obvious in terms of meaning (external heterogeneity) while the codes that have similar associations were placed into the same

category (internal homogeneity). Table 6 indicates the codes and categories developed for each case in this research. More particularly, how language testing governs instructional dynamics such as in-class assessment, learning outcomes and school community is mainly described with the codes obtained from EFL teachers' interviews while the codes generated from Field Experts' interviews were used to figure out the triggering factors. However, in some categories, both interview responses are presented interchangeably or jointly to support the data-driven relationships. In qualitative data analysis, another essential step is to organize how to display the data and visualize the codes and categories by indicating relationships. Miles and Huberman (2004) list a number of matrices, charts and taxonomies that are shaped on the basis of the research aims. Since the purpose of this study is to investigate how washback exists in language teaching and highlight the test effects and triggering forces, the effect matrices were adopted as the main format of data displaying. In addition to depicting the effects and influential factors with causal networks, the raw data that are translated into English were provided with reference to the target codes and categories. The samples of both translated and raw data are given in Appendix C. While most of the data displays are the products of within-case analysis, cross-case analysis results were demonstrated especially in the part entitled "Exam-Oriented Learning Habits" that show long-lasting effect of USE-EE. After the presentation of the data, codes, and categories, the study results were explained in light of the target cases and interpreted by addressing research questions.

Reliability and Validity

Reliability and validity are two fundamental requirements that need to be addressed and ensured throughout the research. In qualitative methodology, validity is related to the verifications of the research data and results with some basic methods, and reliability is concerned about the consistency of the results when the study is repeated in a similar context (Creswell, 2014; Yıldırım & Şimşek, 2013). In case studies, there are four types of reliability and validity: construct validity, internal validity, external validity and reliability (Yin, 2013). Respectively, construct validity is about how the research findings are linked to the research constructs; internal validity is concerned with the appropriateness of the causal relationships and inferences to the qualitative data; external validity focuses on whether the findings are generalizable to other situations or not; and reliability deals with ensuring the same results when a similar study is conducted in a different context.

Table 6

Cases, Codes, and Categories (continued)

CASE 1: PRIMARY EDUCATION (N=24)					
<i>Teaching Methodology</i>	<i>Teacher-Based Assessment</i>		<i>Mock Examinations</i>	<i>Stakeholders Perceptions and Expectations</i>	<i>Teacher-Related Factors</i>
Teaching practices	Evidence on language learning		Feedback on language learning	Use of mock examinations	Teacher Beliefs*
Curricular complexity	Exam anxiety		Misconceptions on learning	Demand on language tests*	Teaching Purposes
Teaching duration	Exam-oriented teaching			Parental expectations	
	Motivation			Parents' ambitions	
	Assessment practices			Upcoming centralized exams	
	Strategy development			Upper levels of education	
	Use of language tests			Institutional prestige	
	Examination regulation				
	Predictions on further success				
	Examination regulation				
CASE 2: LOWER SECONDARY EDUCATION (N=38)					
<i>Teacher-Based Assessment</i>	<i>Mock Examinations</i>		<i>High-Stakes Testing</i>	<i>Stakeholders Perceptions and Expectations</i>	<i>Teacher-Related Factors</i>
Exam anxiety	Accountability	Teaching practices*	Exam couching	Medium of teaching	Teacher identity
Further success	Feedback on learning	Test strategies	Exam awareness	Compliance with examination regulation	Educational background*
Feedback on learning	Content specifications	Assessment practices*	Feedback	Expectations about test-oriented teaching	Pre-service education*
Evidence on achievement	Predictions on achievement	Alternative assessment	Teacher motivation	Success in USE-EE	Assessment literacy*

Examination regulation	Test skills and strategies	Supplementary courses	Learning habits	Previous learning habits
Teaching Duration		Mock examinations	Teacher designation	Test materials
Diploma score		Awarding	Institutional comparisons*	Test-oriented assessment* language
		Socio-economic concerns*		Language assessment policies*

CASE 3: UPPER SECONDARY EDUCATION (N=44)

<i>Teacher-Based Assessment</i>		<i>High-Stakes Testing</i>	<i>Stakeholders Perceptions and Expectations</i>	<i>Teacher-Related Factors</i>
Promoting spoken production	Lerner motivation	Practicality*	Low difficulty of assessment	Assessment literacy*
Feedback on learning	Teacher motivation	Testing facilities*	Use of test items	Teaching practices
Teacher collaboration	High-stakes test specifications*	Reliability*	Employability policies*	Testing practices
Exam anxiety	Attainment of language skills*	Rater-training*	Socio-cultural dimensions*	Teaching beliefs
Excitement	Teacher designation	Objectivity*	Socio-economic factors	Pre-service education
Learning habits	Test-oriented teaching	Accountability*	Institutional competition	In-service education*
Examination regulation	Test-oriented assessment	Transparency*	Knowledge of testing	
Diploma score	Hidden curriculum	Mismatch between testing and curriculum*	Sense of achievement*	
Evidence on success	Test strategies	Teacher practices	Evidence on success*	
	Mock examinations	Testing practices		

This table mostly includes the codes deduced from EFL teachers' responses to basically explore the influence of low-stakes and high-stakes testing on language instruction. Some teacher responses also shed light upon why these effects are triggered.

*These codes are derived from Field Experts' responses to investigate the forces that interplay with the washback effects in primary and secondary education.

In order to meet the requirements of these four aspects of the validity and reliability, it is recommended to use different techniques. In this study, the following ones were used:

1. *Chain of evidence* was established for construct validity.
2. *Member checking* was used for internal validity.
3. *Thick descriptions* were provided for external validity.
4. *Case-study database* was prepared for reliability.
5. *Inter-coder reliability* was determined to minimize the researcher bias.

To be more specific, *chain of evidence* requires indicating the data for making inferences about the research questions; so, all the codes and categories presented here were supported with the rich data that are transcribed and translated. *Member checking* was performed with the participants during interviews in data collection. The researcher repeated the participants' responses during interview and asked for their confirmation about the appropriateness of the inferences. Two examples of member checking are given below:

EFL Teacher Interview:

Researcher: *You think that it depends on teacher knowledge and skills.*

LST-1: *Exactly, I think that it depends on teacher knowledge and skills.*

Field Expert Interview:

Field Expert 6: *We have to consider supply and demand. When it comes to the 8th grade, you are [expected] to conduct exam-oriented teaching. Language teacher says "There is nothing to do because parents have this demand." Therefore, there is a policy in use: hidden policy.*

Researcher: *You have touched upon important points. First you have mentioned the supply and demand issue. There are parental demands [expectations]. They expect teachers to practice more tests.*

Throughout all the interviews conducted with the participants, member checking was managed as shown in the examples above. As another technique used to ensure validity, *Thick descriptions* were presented by giving rich information about how to collect and analyze the data. Finally, a *case-study database* including transcriptions, raw and translated data, initial coding, second-level codes and categories was prepared to carry out the systematic realization of the research steps. Finally, in order to ensure the reliability of the coding procedure, *inter-coder reliability* was calculated. Firstly, the coded data were sampled. Out of 207 codes that

are generated in the initial coding process, 50 codes were selectively sampled from all three cases (The percentage of sampling in each case is 24-26%). It is suggested that sampling percentage should not be below .10 (The sample size in qualitative research is specified as 35% of all the recorded and transcribed data. Neuendorf (2002) claims that at least 50 different pieces of data should be coded by a different coder in order for the researchers to understand the agreement levels. Following this sampling procedure, inter-coder reliability form was prepared. In this form, the research aims, constructs, questions and the cases of the retrieved data were presented to the second coder who is a language teaching expert and also knowledgeable about the research process. Out of 50 codes, 46 agreements were ensured. The five codes that two codes have not agreed on are parental perspectives (the second coding is more comprehensive), teacher behaviours, assessing speaking skills (the second coding is more specific) and accountability (the first coding is more comprehensive). This ratio is equal to 92.5% of agreement (Each code is 2.5% of weight). Miles and Huberman (1994) state that inter-coder reliability should be .80 and above. It means that coding procedure has a high level of reliability. This reliable coding functions as a basis for drawing valid conclusions and interpretations.

To summarize, the current study has employed different techniques to make sure the reliability and validity of the qualitative research. Validity was ensured through presentation of the results supported and driven by the data (chain of evidence), rich descriptions of the case study steps and components (thick descriptions) and validation of the interviewees' responses. As for the reliability, a database was generated and organized in a way of including all the documents and information used in the study. In addition, inter-coder reliability was calculated by examining the agreement levels between two different coding units.



CHAPTER IV

FINDINGS

This section displays the findings obtained as a result of qualitative data analysis. The findings are presented within three cases: primary, lower secondary and upper secondary education. In each level of education, the effects, impacts and consequences of both low-stakes and high-stakes language testing that are the units of analysis are illustrated in regard to language teaching, learning, testing and stakeholders. Following that, macro- and micro-level forces that govern the influence of both language testing types are highlighted with respect to the dynamics of EFL teaching and learning setting. EFL teacher responses are mainly used to scrutinize the influence of language testing, and Field Expert responses throw light upon the triggering factors and interaction of these factors with washback effects.

Case 1: Primary Education

In this part, it is attempted to demonstrate how low-stakes assessment affects language teaching, learning, assessment and stakeholders in primary education. Besides, the triggers that exert these effects on EFL dynamics are presented in relation to the specificity and intensity of washback effect. The following two subsections portray the influence of written exams, mock tests and practice tests on language teaching and testing practices and how these effects are shaped by political, social, economic and professional triggers. At the core of these triggers, third subsection centralizes on how language teachers take a role in mediating washback effects.

Washback Effects of Formal Written Examinations and Triggering Factors

Methodological Shift in Language Assessment

English language is assessed with different techniques, tools and methods in the 2nd, 3rd and 4th grades as summarized in Figure 5. Especially in the 2nd and 3rd grades, there are no official written exams as PLST-2 and PLST-3 explained. Instead, teachers use assignments, projects, portfolios in order to monitor their students' learning development. More particularly, role-plays, songs, and games are utilized for both learning and monitoring purposes. On the other hand, formal written exams are administered in the 4th grades and onwards.

“There are no written exams in the 2nd and 3rd grades at primary schools. Assessment and evaluation in these grades are based on the assignments and projects. We begin administering written exams in the 4th grade and we do not conduct exam-oriented teaching in the 2nd and 3rd grades.” [PLST 3]

When PLST-2 expresses her assessment practices, she underlines that the primary school learners are assessed in accordance with the alternative ways of assessment that are instructed in pre-service education. For instance, she conducts task-based assessment, portfolio assessment, and performance-based assessment. In the formal examination in the 4th grade, she avoids using multiple choice items to measure target objectives.

“There are some differences between primary and lower secondary education. We are not expected to administer written exams [in the 2nd and 3rd grades] at primary schools but there should be written documents of the exams at lower secondary schools. At primary schools, as we [practiced] at university, we employ some tasks and get them to prepare learning portfolios. In this way, they can see what they have learnt. Besides, we carry out performance-based assessments. We use drama, role-plays and songs that are important for us in assessment and evaluation. ... We do not prefer using tests in written exams [in the 4th grade]. They are based on matching and comprehension.” [PLST 2]

However, PST-3 highlights the use of multiple choice items in the 4th graders' written exams. This points out a discrepancy in teacher assessment practices. In view of the coverage of the language skills in the assessment of the 4th grades, formal exams mainly focus on reading and writing skills with multiple choice, gap-filling and matching item types as PST-3 stated. On the other hand, oral skills (i.e. speaking and listening) are not directly and explicitly tested in the formal examination. Instead, in-class observation, as one of the alternatives in assessment, is conducted to assess listening and speaking skills during the lesson in an

integrated way. PST-3 also asserts that technological infrastructure and materials helped them both teach and assess the listening skill in the 4th grade.

“I do not measure all the skills in the written exams in the 4th grade. Although we put emphasis on speaking skills in the class, I have not measured this skill. I measure reading comprehension and writing skills through multiple choice items, gap-filling and matching.” [PST 3]

“We observe those skills [listening and speaking] in the class. There are [technological] facilities in the school. Smart-boards are an advantage for us. There are interactive course books that are user-friendly now. We can play an audio recording easily and do the activities related to this recording. [PST 3]

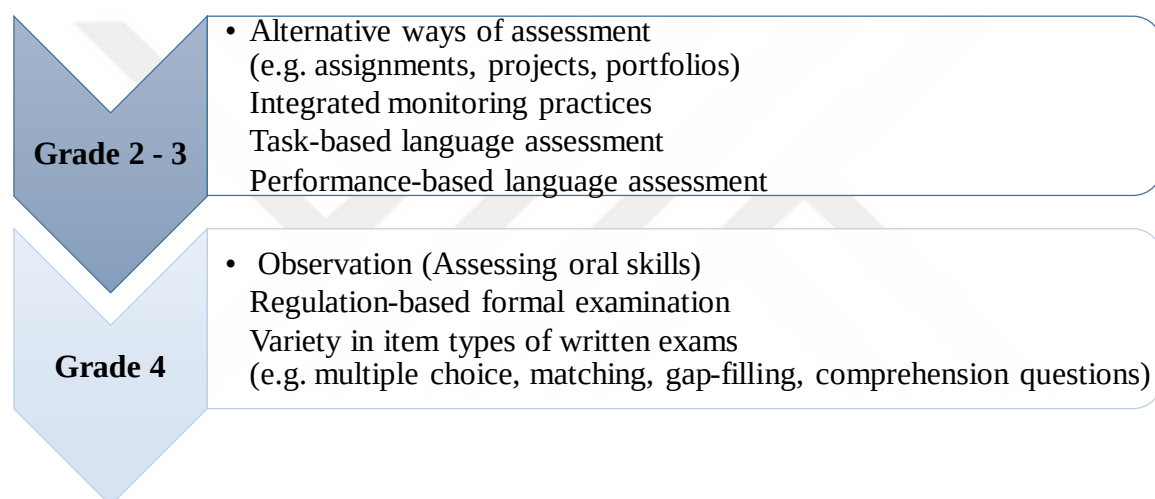


Figure 5. Framework for teacher-based assessment at primary state schools

It shows that the 2nd and 3rd graders are assessed in consistence with instructional practices. EFL teachers also use alternative ways to keep track of their language development through portfolios, projects and tasks. In the 4th grade, there are similar practices but the emphasis is on mainly written exams that chiefly measure reading and writing skills with matching, gap-filling and multiple choice items. Besides, observation is used to assess oral skills that are not included in formal examination.

The main reason and primary force for this methodological shift between the 2nd-3rd grades and 4th grade is viewed as examination regulation itself as can be concluded from PLST-2 and PLST-3’ responses. Whereas there is no formal examination in the 2nd and 3rd grades, learners face the implementation of written exams first in the 4th grade.

Exam-Oriented Language Teaching

The existence of the official written examination in the 4th grade forces PST-2 to review the exam content and course coverage. He believes in teaching each detail that is given in the curriculum because he has to assess these details with two formal exams. He has no chance to disregard the curricular content that is to be measured in the exams. For that reason, his concerns on teaching for the sake of the exams overweight in this grade as pointed out in the PST-2 response:

“While we enjoy language learning in the 2nd and 3rd grades, we feel like in the 4th grade that these students will take exams, and we must not assess what we do not teach in the 4th grade. For example, we are going to administer two exams in the fall term; so we have to cover the first 4-5 units. In this way, I teach what I plan to assess. Thus, I give more weight to the exams and teach in line with them.” [PST 2]

Not only does the exam coverage influence teaching content, but also the complexity of the topics and transition to higher grades accompany the exams in presenting the precise content by allocating less time to enjoyable tasks. PST-4 states that there is an instructional shift from the 2nd and 3rd grades to 4th grades in terms of language production. Also, she explains that the 4th graders are getting prepared for lower secondary education. In order for them not to face any difficulties in upper grades, she allots less time for the activities like games in the 4th grade.

“Content is getting more complex [in the 4th grade]. Whereas our teaching requires more word-level production in the first two years [the 2nd and 3rd grade], the students are expected to make sentences and have more communication in the 4th grade. That is why we spend less time for games and allocate more time for teaching the content so that the students cannot have difficulty in language learning in the 5th grade.” [PST 4]

While PST-2 emphasizes the influence of written exams on the course content in the 4th grade and also complexity of the content affects the learning practices as shown in PST-4 response, PST-3 and PLST-3 evidently state that they do not carry out exam-oriented teaching even if it is necessary to administer two formal exams. This finding highlights the divergence again between language teachers in terms of adopting an exam-oriented perspective.

“We have more drama activities in the 4th grade course books and we use them. As I said before, our teaching is not based on the exams. It is not exam-oriented even in the 4th grades.” [PST 3]

“We conduct written exams but we do it in order to comply with the formality. They do not affect the lesson as much as the exam affects the 8th graders. Our main focus is on the process, not on the exam. The students get the outcome of this process in the exams. We do not aim to score them lower or higher; instead, we prepare the exams to motivate them.” [PLST 3]

Even though the content is more complex and it is required to conduct formal examination in the 4th grade, PST-3 does not adopt an exam-oriented teaching perspective. She employs the communicative tasks provided in the course books. Similarly, PLST 3 carries out written examinations because of the official regulations and does not put stress on the exam scores. Rather, she highlights the learning process and assess her students with motivating purposes. So, it is seen here that language assessment in the 4th grade is performed to motivate the students to learn English language.

Feedback on Language Learning

In addition to this motivating aspect of the formal examination, PST-2 utilizes the exam results to define learning strengths and weaknesses. By examining what has been learnt and considering what should be done, PST-2 plans new strategies for further academic years.

“While evaluating the general achievement, we make some decisions for the following year. For example, we should give more weight to some topics. Then we try to do in this way. Simply, we attempt to recognize our weaknesses.” [PST 2]

Hence, it is concluded that exam results pave the way for analyzing the current learning setting and making decisions on teaching to improve the quality of education.

Exam Anxiety

Since there is no formal examination in the 2nd and 3rd grades and the students’ purpose is to have fun in the lesson and enjoy learning English, they are not aware of being monitored and evaluated as “very good”, “good” or “needs development”. Their awareness on learning rather than exams and no explicit and official implementation of written exams in these grades result in less anxiety and negative feelings for language assessment as PLST-1 said.

“There is no exam anxiety at primary schools. The students do not care whether their overall grade will be “good” or “needs development”. As we teach English through songs, games and tasks, they just think that it is an activity-based lesson and they are not aware of it.” [PLST 1]

PLST-2 also touches upon her motive not to create a competitive learning setting for young learners. Without the sense of being measured and evaluated, the 2nd and 3rd graders enjoy testing their hypotheses in speaking and making mistakes through active involvement. She hesitates about the introduction of explicit language assessment into these grades for the fear that this engaging environment will be damaged because of examination-based competition.

“We do not want our students to get into the mood of being measured and evaluated. They are not aware of being tested. They are active participants in the class, try to speak and make lots of mistakes. If I tell them that I am going to administer exams, I know that this atmosphere will be demolished. [PLST 2]

On the other hand, through the 4th grades, the students get anxious about the implementation of formal examination. Even if the teacher, PST-2, tries to relieve them, the students are still afraid of getting low scores in the exams.

“The 2nd and 3rd graders do not have any anxiety about the lesson or exams. Since there is no such an anxiety, the students learn the language easily and enjoy it. However, some of the 3rd graders state that they enjoy a lot this year but they will have exams in the following year. I just tell them that there will be exams [next year] but they should not be afraid of them. Even so, they have the fear of getting low scores. That is, we get them to feel anxious” [PST 2]

When there is no fear or anxiety of examination in a classroom, the students enjoy language teaching; on the contrary, the fact that the students feel anxious or nervous about the low scoring in the written exams influences learning negatively.

Parental Expectations

One of the reasons why the students feel anxious about the scoring in the exams could be their parents' expectations. As PST-3 puts forward, parents are mostly product-oriented. Most of them evaluate their children's performance according to the exam results rather than learning process. Their expectations about getting higher scores in the examination push the students to spend excessive amount of time and energy to think about the examination. Instead of focusing on language learning and enjoying the activities, they become more curious about their exam results.

“When I was appointed to this school, they told me that parents here gave more importance to the exam scores. I really observe that there are some parents that focus on the outcome rather

than the process. The parents not interested in learning process ask me to show the exam documents if the exam results are not as they have expected. They claim that their children should have higher scores. [PST 3]

Some other parents may approach the phenomenon of examination with a different perspective. The primary education is the beginning level of academic career, and these parents assume that the academic scores will estimate their future academic achievement in the upper grades. PST-3 divides the parents into two groups; the parents in the first group regard the exam scores as the basis for further learning and make some judgments on their children's performance. In the other group, parents' ambitions affect the student behaviours. In this group, parents have higher expectations from their children. For instance, they want them to be successful all the time and get higher marks from the exams. It causes anxiety among the children and leads them to put more emphasis on learning outcomes.

"Some parents think that these scores serve as the basis for their children's further success. Some others, let's say who have personal ambitions, think that their children are very successful, and they do not accept the failure [the score below a specific level]. They think that their children must always be successful. I believe that there are two groups, one of which is the parents who have biggest personal ambitions, and other one of which is the parents who think that the scores are important for the following years since they could be a learning habit.".

[PST 3]

PST-3 responses elucidate that parental behaviors are the main sources that shape the learners' attitudes towards language assessment as pointed out in Figure 6. If the parents expect their children to have high scores that are accepted as the evidence for learning and further success, the learners will focus on getting only good grades rather than enjoying language learning.

Figure 6 displays the relations between formal examination effects, consequences and influential factors. To be specific, written exams bring about an assessment shift towards the 4th grade. This shift is triggered by not only official policies but also curricular content and time. Also, written exams drive some language teachers to adopt an exam-oriented stance in teaching. In contrast, some teachers do not base their instruction on written exam content or methodology. By the same token, some teachers do not allow the stakeholders to prepare a competitive learning setting that may increase the level of anxiety. This finding indicates that teachers mediate the washback effects in primary classrooms. In terms of positive value

of washback, EFL teachers make use of exam results in order to revise their teaching procedures and practices.

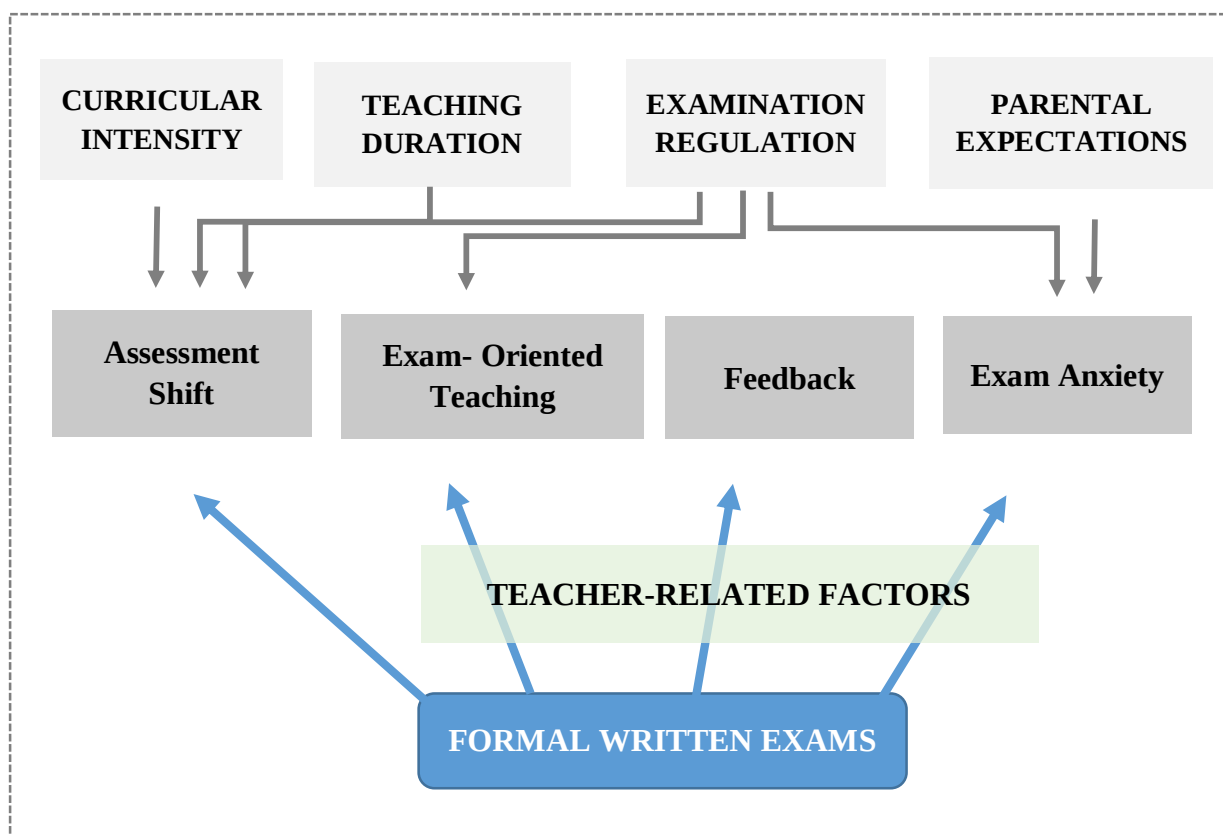


Figure 6. Washback effects and triggering factors in primary education

Use of Mock Exams and Practice Tests

Mock Examination Washback and Triggering Factors

In addition to teacher-based assessment including monitoring practices, alternative ways of assessment and formal examination, there is another type of examination that is commonly used at primary schools. It is the mock examination with large-scale implementation. This exam consists of multiple choice items and may be employed in different learning settings. As PST-4 underlines, this exam type is not appropriate to the curricular background of teaching young learners. The primary school students have enjoyable time in learning English, and multiple choice items are not presented to them in these grades, especially in the 2nd and 3rd grades. Nevertheless, these exams are frequently used in this level of education due to some factors exemplified in the following paragraphs. PST-4 accentuates that young learners are not familiar with this kind of testing and, at the end of testing, they

have misconceptions about language learning. Furthermore, EFL teachers' teaching practices and coverage of the curricular content are discussed by the stakeholders like classroom teachers and parents. When the students fail in answering all the multiple choice items correctly, the parents make EFL teachers responsible for the mock exam results.

"As we have enjoyable time in English language lessons [in the 2nd grades] and students are not familiar with the examination settings, the 2nd graders feel astonished when they take the mock exams that include multiple choice items. We attempt to help the students be exposed to the language; [so]they do not know how to write in English. For that reason, they are puzzled [when they take mock exams] and they think that they have not learnt. Then the classroom teacher makes judgments on whether I have taught or not. Also, parents make me accountable for the low ranking of their children in the class. We face this challenge because of the parents' anxiety." [PST 4]

In a similar vein, PST-3 draws attention to the content and construct validity of mock exams and explains that the lack of validity in these exams causes parental concerns. In consonance with PST-4's response, PST-3 mentions the accountability issue at the end of such examinations and also highlights that the stakeholders may make comparisons among the subjects in line with the exam results. Even though English language is compared to other disciplines in terms of the number of correct answers, PST-3 examines the results and focuses on learning strengths and weaknesses.

"Mock exams may include the questions that are not included in the curriculum. In this case, the parents are getting anxious. We relieve them by explaining that there is no such a topic in the curriculum. What's more, [one day] classroom teachers expressed that all the students were successful in the mock sub-tests except for English test in which they had two correct answers out of five questions. They may compare the subjects but we do not compare the grades. We just examine our learning weaknesses." [PST 3]

As explained above, mock exams are used in the primary education, and they have some consequences for language learners, language learning and EFL teachers. PST-1 believes that the reasons why this exam type is so commonly used in this level of education and influences EFL dynamics are related to the classroom teachers' concerns about the students' transitions into lower secondary education and availability of a centralized exam at the end of the 8th grade.

“I do not think that use of mock tests is not related to their [classroom teachers] workload. It is related to the system of the centralized exam that exists in our general education. The students will get to know this system in the 5th grade. The classroom teachers have the concerns that [in lower secondary education] the 5th graders’ previous teachers (primary school teachers) will be judged if they do not provide opportunities for the students to take the tests and fill in the answer sheets. They employ these tests and exams in consideration of the centralized exams.”

[PST 1]

Classroom teachers think that their teaching and testing ways will be criticized if the primary school students are not trained about how to answer multiple-choice questions. In this regard, the transition to lower secondary education triggers the use of tests and mock exams in this level of education. Likewise, PST-4 comments on the forces that prompt the popularity of mock exams at the primary level. Similar to the classroom teachers’ concerns about the transition to lower secondary education, the parents feel anxious about the placement of their children into “good schools”. Good schools offer a promising future to the students and the only way to get education at those schools is to be selected in the large-scale examination.

“Unfortunately they give importance to use of mock exams and the results of these exams. We explained that there should not be mock exams in the 2nd grade in our meetings, and it caused a decrease in the use of those exams but they are still important for the 4th graders. They are important because the parents are anxious about the placement of their children into ‘good schools’ with good teachers and high level of achievement, the principals get prestige according to the achievement level of the schools, and the only way to achieve this [to announce that a school is good and successful] is through measurement and evaluation. To be honest, the school grades are not interpreted [in our school] but the administrators arrange meetings with the teachers about the results of the mock exams in order to improve the scores.” **[PST 4]**

Like the parents who are concerned about their children’s academic career and future life, the administrators attach importance to the use of such exams in primary education. PST-4 mentions why mock tests and their results have a pivotal importance for the administrators. The results of mock exams enable the stakeholders to make comparisons among the primary schools, and the high ranking emerged at the end of examination gives prestige to the schools and labels them as the institutions with “high achievement”. Thus, administrators support the use of mock tests at primary schools and encourage the others to improve the test results.

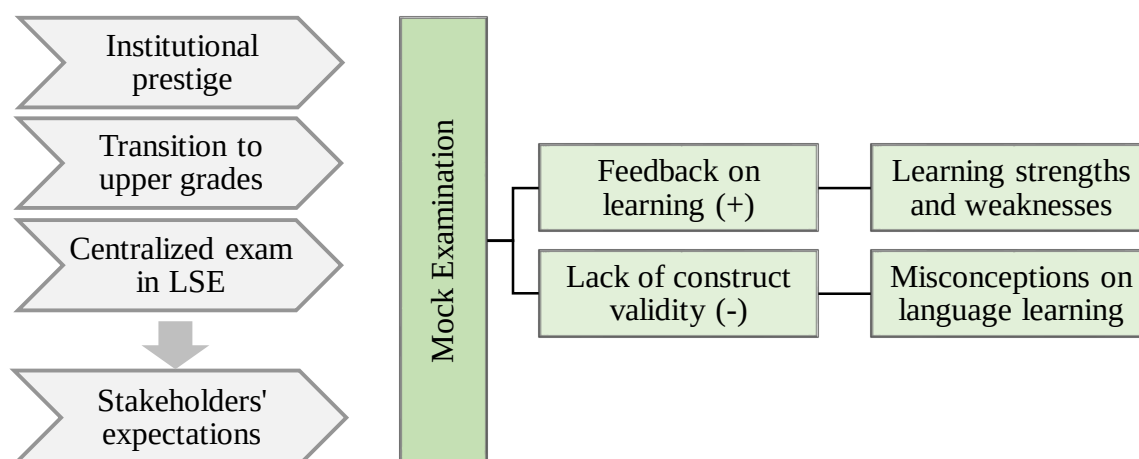


Figure 7. Use of mock tests at primary schools: External forces that mandate the use of mock tests and outcomes for language learning

In the figure presented above, the effects of mock examinations based on multiple-choice items are holistically portrayed. USE-EE, transition to higher levels and institutional dynamics shape the stakeholders' perceptions. These perceptions promote the use of mock examinations. On the one hand, there is a concern about the test specifications of mock tests. It is claimed that out-of-curriculum content is measured in these tests. It directly influences construct and consequential validity. Therefore, some parties may develop misconceptions about language learning and teaching. On the other hand, test results may give feedback on what language objectives are achieved. It is the evident beneficial influence of mock exams on teaching.

Washback Effects of Practice Tests and Triggering Factors

The concerns about further academic career and aiming to increase school prestige result in the use of mock exams in the primary education. In addition, the parents' perceptions on teaching and testing put language teachers under pressure in terms of use of language practice tests in the classes. PST-1 shares his observation about how parents interpret the effective learning. The parent supports the traditional way of teaching and disapproves the use of communicative tasks like games or songs. According to this parent, the students should take more language tests and do more writing practices.

"I have witnessed this week that one of the parents would like to change her child's [language] teacher because this teacher brings games to the classroom, the students have fun and laugh.

She is not sure about whether the teacher lectures or not. The parents' perceptions on learning are that teachers employ tests, want the students to write something, assign them [some] homework etc. For this reason, she wants her child to be trained by another teacher. They think that he/she is not a good teacher because of using songs, tasks or draw-and-paint activities."

[PST 1]

Considering the parents' perceptions on traditional teaching, Field Expert 3 discusses how these perceptions are shaped in the society. He puts forward that testing system itself directly determines what is valuable in the society. If you test something, it becomes important. When the testing is conducted with multiple choice items rather than authentic or communicative tasks, parents demand the use of test in the classroom.

"Which is all about what they have been exposed to, which is all about to what extent they are expected to use traditional or modern approaches obviously. Perceptions that the society might have might be very much related to examinations if you know student selection and placement the center or selection examination, if they highly give importance to tests, the society will believe that it is the test that matters only. Society beliefs or other things are presumably influenced by what is essential, what is mandatory, what is crucial in this regard. [Field Expert 3]

Another factor that frames the parental perceptions is the high-stakes role of testing in the employment of their children. Large-scale tests are administered across the country in different stages of education with the purposes of selection and placement. At the end of lower secondary education, the students find themselves in competition to be placed into the upper secondary schools which are appropriate to their interests and further academic plans. At the end of upper secondary education, they again put effort to pass the university exam and to be placed into the departments in which they will gain mastery. These departments will provide the students with employability opportunities in order to get a prestigious job and earn money. In a nutshell, large-scale tests directly affect the employability and socio-economic status of the family. That is why the parents attach importance to the exams and ask the teachers to practice the tests as pointed out in PST-1's response.

"One of the important reasons why there are such perspectives on testing and evaluation is that parents think these exams and higher exam scores are significant in terms of getting education at the schools with good quality and then getting a qualified job. ... In the current system, the parents think that if their children do not have education at an Anatolian or science upper secondary school with good quality, there is no possibility for them to become a doctor, an

engineer or a judge. As the students are placed into these schools with test scores and diploma scores, testing and evaluation gain more importance [there].” [PST 1]

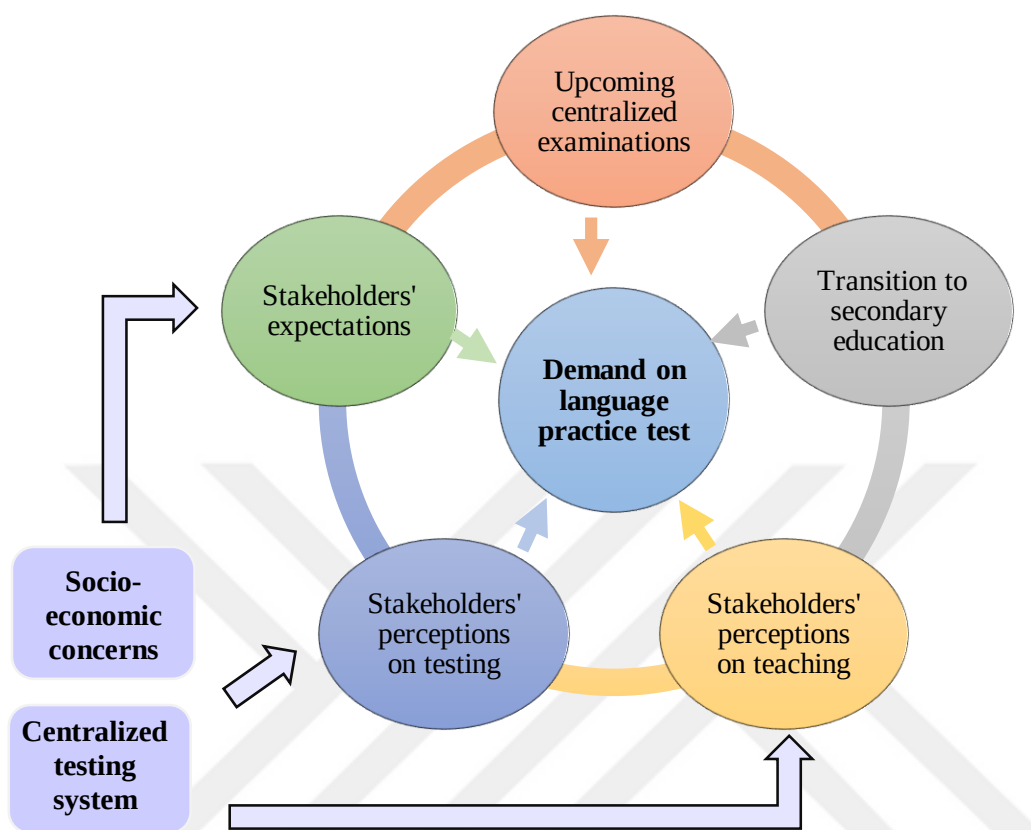


Figure 8. Influential factors that increase the demand on the use of multiple-choice items included in language tests

Besides the large-scale tests that are administered at the end of lower and upper secondary education, primary school students get prepared for the scholarship exams conducted by the private schools in lower secondary education. Since the specifications of these exams are based on multiple-choice items, the students want to make practices to take tests. PST-3 emphasizes that especially the students whose parents have high awareness of such exams have more demand on the use of tests as the instruction and practice tools.

“There are some students who want to take the scholarship exams of the private lower secondary schools. The students whose parents have awareness [about these exams] ask us to employ more tests since they plan to take these exams in the following year.” [PST 3]

Figure 8 illustrates the circle of triggers that facilitate the realization of washback effect: use of test items in learning setting. The macro-level outer forces, socio-economic concerns and high-stakes testing policies direct the stakeholders’ language teaching and testing

perspectives and expectations. They focus attention on only subject matters and techniques that are measured in these high-stakes tests because this deliberate attempt provides a lever for socio-economic changes. Additionally, upcoming USE-EE and taken-for-granted perceptions about language teaching in lower secondary education drive teachers to put emphasis on use of language practice tests.

While some stakeholders have a test-oriented approach, PST-3 rejects this point of view and uses the test items only in order to practice language units. When the students want her to practice language tests in the primary grades, she meets this demand in the supplementary hours of teaching (i.e. additional language courses given at the end of the fixed schedule), not in the regular teaching hours. She clearly states that teaching is not test-oriented even if they use the tests in the class with the instructional purposes.

“We have supplementary courses [in addition to the fixed schedule] and there is much demand on English language lessons. We use more tests in these courses. However, in our fixed schedule, our focus is not on use of tests; instead, we employ dialogues, games and tasks.... We have no such an understanding [of test-oriented teaching] but it reinforces their practice. Our teaching is based on vocabulary like colours, numbers rather than grammar at primary schools. We do not teach advanced grammar but when they learn “do”, “am/is/are” or “can”, test is the appropriate tool for them to find the answers quickly. Use of test helps them give quick responses in the question and answer activities or role-plays but I cannot say that we teach on the basis of tests in even the 4th grades. We teach in line with the curricular objectives. If the example in the interactive course books is not enough for the students to understand, we support their learning with some online applications that have free access. [PST 3]

She accentuates that the advanced and complex grammar is not taught in the primary level; rather, there are basic structures like auxiliary verbs or modals. Multiple choice questions help them recall the structures and then lead them to find the answers in a practical way. In other words, test is a tool for supporting learning but it does not mean that test mandates the learning. PST-3 never adopts test-oriented teaching in spite of the upcoming large-scale centralized exams and stakeholders’ demand on practicing language items for mock and scholarship examinations. She aims to implement the curricular principles in the lessons. As it is obvious, there are teacher-related differences in situating language teaching on a test- or exam-oriented basis. These differences mediate the intensity and directions of washback

effects in primary education. Such teacher-related factors will be illustrated in the following section.

Teacher Beliefs and Missions

In the previous sections, some influential factors that trigger the use of different types of exams, test practices and washback effects were presented. Classroom teachers' concerns, formal examination expected in the regulation, stakeholders' expectations and perceptions interfere with EFL teachers' teaching and testing practices. Under the pressure of these forces, some EFL teachers may focus attention on exam-oriented teaching and common use of language tests and mock exams. On other hand, there are some EFL teachers who do not allow other parties to govern their instructions like PST-1. Typical characteristics of these teachers are that they are self-confident and they trust their knowledge, skills and experiences. For example, PST-1 tells them the stakeholders that he is the expert of language teaching, and he is going to teach in line with his truths and beliefs.

"Is it my luck, my identity or personality, I do not know but I have my own truths. I do not ignore what others say; of course, if they are beneficial for my instruction, I take them into account. At the beginning of the term, I talk to the classroom teachers and parents about testing and evaluation and tell them that I am the teacher of this lesson, the expert who gets training and has experiences in this field. I explain what I consider appropriate to this age level and what is not. 'These are my methods and techniques. You do not need to tell me about what other teachers do. Thus, do not ask me why we do not use tests and buy supplementary books.' I make this explanation to the class teachers and parents at the beginning. So, I never employ more exams in primary education.". [PST 1]

According to PST-3's beliefs, a language teacher is the main figure in teaching process, and therefore they should not be driven by other stakeholders' conceptions and demands. He puts his beliefs into practice in his instructional setting. Another dynamic that shapes the instructional practices is teaching purpose. The curricular goal is simply to provide age-appropriate learning experiences with young learners and get them to have positive emotions for English language. PST-3 aims to act accordingly in consideration of these goals. She tries to avoid pushing the students to compete with each other.

"As teachers, our primary aim is to realize that these students are not in the examination competition.... My aim here is to teach basic English to the students and help them develop

positive feelings towards English language because if the students close their doors [to learning], it is really hard to reopen it at lower secondary schools. [PST 3]

She also claims that she would never change the ways of teaching even if there could be a possibility to use the formal examination in the lower grades (2-3). It shows that her educational philosophy leads her to plan her own teaching and testing procedures.

“Our primary purpose is to help the students develop positive feelings towards English language lesson. Since there is no high-stakes examination in the 4th grade like Anatolian lower secondary school exam [LGS], there is no competitive atmosphere in the 4th grade. Even if there were written exams in the 2nd and 3rd grades, we would teach in the same way through the tasks and games.” [PST 3]

On the other side, PST-3 indicates that there is no large-scale centralized test implementation in primary education, and therefore there is no competition among the students in the class or school. Hence, this response shows that the implementation of high-stake examination in primary education might change how teacher-related factors mediate the washback effects.

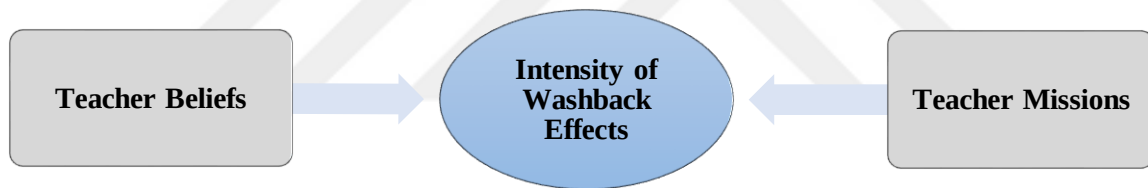


Figure 9. Teacher-related factors that influence the intensity of washback effects

As pointed out in Figure 9, especially in primary schools where no high-stakes testing is introduced, teacher beliefs and missions function are the mediating factors that lessen or increase the intensity of washback.

Case 2: Lower Secondary Education

In this section, washback effects of low-stakes and high-stakes language assessment are described in terms of the influence on language teaching, learning, testing, learners and teachers. What follows scrutinizes macro- and micro-triggers that facilitate these effects. Teacher responses shed light on elaborating on washback effects of language assessment whereas expert views provide insights about what influential factors are hidden behind the

washback effects. These influential factors are also expanded and supported with teacher perspectives. Accordingly, there are three main subsections here: (1) low-stakes language assessment washback and triggering factors, (2) high-stakes language testing washback and (3) triggering factors that exert the washback effects of high-stakes language testing. The first part discusses how teacher-based assessment and mock tests influence language teaching, learning, learners and teachers. Then it sheds light upon the factors that bring about these influences. In the second part, the effects of large-scale centralized test are illustrated with respect to language teaching, learning, testing and policy-making. Finally, the third part uncovers the dynamics that facilitate the washback effects and characterize them macro-level and micro-level triggers.

Washback Effects of Low-Stakes Assessment and Triggering Factors

This part basically unfolds the washback effects of low-stakes assessments: teacher-based assessment and mock examination in lower secondary education. The first section illustrates how and why anxiety increases at the end of examinations. It also reveals the role of assessment in predicting the success, getting feedback and evidence on learning. The second part centralizes upon the use of mock examination and demonstrates how such tests determine teaching content and study skill and how accountable EFL teachers become at the end of test administration. On the other hand, it explains to what extent the results of mock examination serve as a basis for revision of teaching practices and procedures. In the second part, a symbiotic relationship is indicated between mock test washback and trigger. Holistically, washback effects and triggers of low-stakes testing in lower secondary education are presented in this part.

Teacher-Based Assessment Washback and Triggering Factors

Classroom-based assessment including both formative and summative procedures brings about some effects on language learners, learning and teaching. Primarily, it causes anxiety among the lower secondary students. Those who aim to achieve higher scores avoid making mistakes in their assignments or exams. This attitude leads them to convince other parties to complete the assessment tasks on behalf of themselves as depicted in LST-1's response. She believes that these students try to make mistakes if there is no fear or anxiety of examination.

PLST-2 supports the idea that the students have exam anxiety in this level of education and links this situation to the official regulation.

“I do not want them to have their homework done by others. They push the parents or others [to do their homework] in order to get higher scores. What if there is no scoring or exam anxiety? They will not have any difficulties when they make mistakes.” [LST-1]

“The rationale for written examination changes in lower secondary education. Exam anxiety increases there.” [PLST-2]

LST-V-2 and LST-3 also put emphasis on the influence of diploma scores on transition to upper secondary education as another triggering factor in addition to the requirements and principles of examination regulation. This diploma score is not only used in calculation of USE-EE scores for transition but also influential in the student recruitments for upper secondary schools without any centralized exam.

“The students’ diploma score plays a significant role in transition to upper secondary education. For this reason, it is getting more important in the 8th grades.” [LST-V-2]

“[Our formal examination] does not affect the result as much as the centralized exam does. However, the students [may] apply for the schools that do not recruit them with the centralized exam scores, and those schools consider the diploma scores.” [LST-3]

There is an obvious relationship between transition to upper secondary education and low-stakes examination anxiety that is triggered by high-stakes testing policies. Besides this relationship, teacher-based assessment is viewed as the predictor of further success. PLST-1 asserts that parents make some judgments on their children’s future academic career in line with formal examination results. What’s more, the students take the advantage of high scores in proving their success in the society and get the others’ approval of achievement as emphasized by LST-2. In doing so, they develop a sense of achievement and raise the others’ awareness on how successful they are.

“Parents believe that their children will be unsuccessful [in the centralized exams] when their [teacher-based assessment] exam scores are low.” [PLST-1]

“Students are in attempt to prove themselves. Perhaps they want to get respect and approval of their parents. They try to tell others that they achieve something.” [LST-2]

With respect to instructional setting, EFL teachers draw on their assessment results in getting feedback on learning gains as discussed in the responses of LST-V-1 and LST-V-2.

Accordingly, learning strengths and weaknesses are analyzed during or at the end of the assessment process. In line with the analysis outputs, teaching-related decisions are taken and shared with significant others.

“It is important to us. Obviously, it is written feedback for us about to what extent we have taught and the students have learnt.” [LST-V-1]

“It is a kind of evidence on teacher observation. For example, there is a [good] development in [language] learning in the 7th grade, and exam scores are viewed as the evidence for this development. If there is a failure, we try to inform parents about this failure and expect their support.” [LST-V-2]

Taken together, teacher-based assessment causes exam anxiety that is triggered by its high-stakes impact on upper levels of education. In addition to this affective aspect, the results of teacher-based assessment provide solid evidence for making judgments about the students’ success. These results are also seen as a way of receiving the acceptance from the society about the sense of achievement. From instructional perspectives, EFL teachers evaluate the assessment outcomes and revise their own teaching in consideration of the attainment of learning objectives.

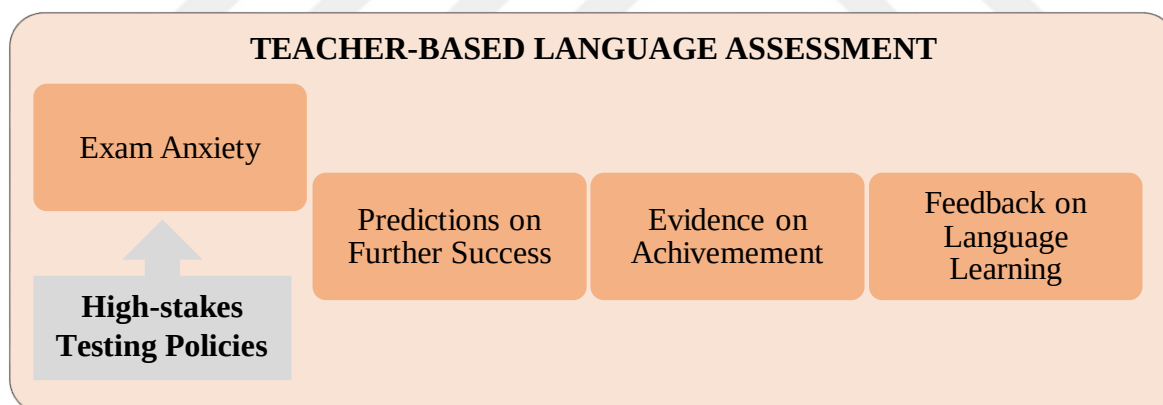


Figure 10. Washback effects of teacher-based assessment on EFL teaching dynamics and triggering force that increases exam anxiety

As seen in Figure 10, there are four outcomes of classroom-based assessment washback: anxiety, prediction, evidence and feedback. Among these factors, examination anxiety is provoked by testing policies of USE-EE.

Mock Examination Washback and Triggering Factors

As the second type of low-stakes assessment, mock exams that are large-scale multiple-choice tests are commonly employed in lower secondary education. These exams do not officially influence the students' overall grade or diploma score. However, it brings about some washback effects. To take an example, PLST 1 comments on how the mock tests influence her way of teaching. She mentions her attempts to introduce mock tests to the students by focusing on the types of items. Evidently, there is an effort to make the students equipped with test-taking skills; therefore, they will perform better in these tests.

"We try to get the students familiarized with the item types in mock tests in order for them to understand the questions [better]." [PLST-1]

"There are sequencing words like firstly, secondly. [Even if] they are not covered in the curriculum; you have to teach them. There is no way. [If you do not teach them] After the students take the mock test, they claim that teacher has not taught the content measured in the test." [LST-V-2]

Besides the familiarization with the item types, LST-V 2 expresses that teaching content also can be shaped by the mock exams. For example, the grammar topic, *sequencing words*, is not included in the ELT curriculum. On the other hand, mock tests may include this topic. [This aspect brings the issue of construct and content validity into questions.] Hence, the coverage of mock exams becomes hidden curriculum that EFL teachers are obliged to teach. Otherwise, it is assumed that teachers do not carry out their teaching tasks as expected.

Besides test alignment and test-taking strategies, accountability is an issue that is largely argued as a result of mock test administration. According to parents, mock test results reflect upon what has been taught and what has been not in the classroom. So, they make language teachers accountable for either failure or success. As expressed in LST 3's response, EFL teachers are questioned about whether they presented the vocabulary measured in mock exams or not.

"Parents tell me that their children have only 3 correct answers out of 10 [multiple-choice] questions in the first mock exam and ask me whether I have taught the vocabulary or not." [LST-3]

"My [teaching] success depends on the mock test results. Parents think that I am a successful teacher when their children have 9 correct answers out of 10 questions." [LST-1]

Whereas this kind of low-stakes assessment shapes teaching content and makes teachers responsible for the results, PLST 1 mentions the formative aspect of mock exams. This response clarifies that the exam results give an overall picture of learning by summarizing the strengths and weaknesses. Within the frame of this picture, EFL teachers make some revisions to promote the achievement of language objectives and support learning with additional practices.

“We get the exam results and analyze them. Based on these results, we revise our teaching. For example, we give more assignments to the students.” [PLST-1]

It is evident that use of mock tests results in the alignment of language teaching to test content. Also, stakeholders develop an understanding about the fulfilment of teaching job on the basis of test results. The main influential reason for administering these tests and allowing them to bring about washback effects is their key role in predicting the students’ further as elucidated in LST-V 2’s response below.

“Mock exams are important to us. They are monthly or weekly administered at some schools. At the end of the examinations, [the stakeholders] make some comparisons and judgments about the students. For example, they focus on which student promises the future or passes the centralized exam.” [LST-V-2]

LST-V 2 discusses that there is a common belief that students who get higher exam scores will pass the university exam. Thus, it is assumed that mock exams predict the student further success.

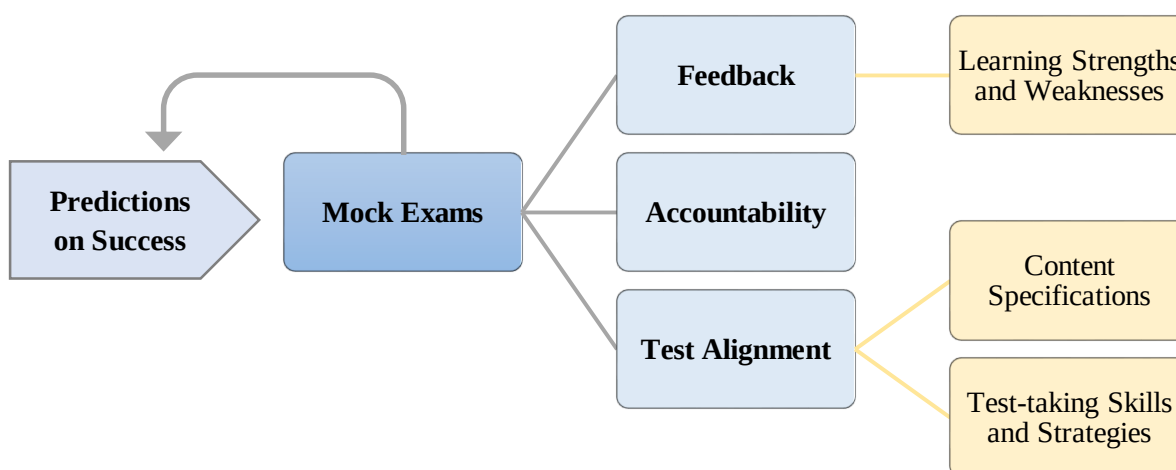


Figure 11. Use of mock exams in lower secondary education, force triggering their use and washback on language learning and testing

Figure 11 visualizes the washback effects of mock exams and the main trigger of these effects. Specifically, mock tests are used to predict how well the students can perform in upcoming centralized exams. This predicting power shapes micro-level institutional policies that require the students to take mock tests regularly. Additionally, there is a symbiotic relationship between mock test washback and predictions. It is both outcome and trigger of the mock examination. As mentioned in the previous arguments and shown in Figure 11, this examination specifies some parts of teaching content, directs the students to practice test strategies and helps teachers understand how effective language teaching is.

Washback Effects of High-Stakes Testing and Triggering Factors

In the previous section, low-stakes language assessment was examined with regard to how it influences language teaching dynamics. This part draws attention to the high-stakes testing: USE-EE. Firstly, the major effects of this large-scale exam will be displayed in relation to language teaching, learning, testing and policy-making. Following that, the underlying factors that have an essential role in exerting these effects will be depicted in consideration of the macro-level (testing policies, and socio-economic forces) and micro-level (stakeholders, institution, teaching and teachers) dynamics.

USE-EE Washback on Language Teaching, Learning and Teachers

Methodological Shift in Language Teaching

A majority of EFL teachers who work in lower secondary education put stress on the existence of methodological shift in language teaching from the 5th and 6th grades to the 7th and 8th grades. For instance, LST 1 conducts a needs analysis in the 5th grade and adopts an eclectic approach including TPR. However, this perspective paves the way for exam-oriented teaching in the 8th grade. Mechanical activities and lexical memorization are the main types of instructional practices. Teaching purpose here is to help the students practice USE-EE with both in-class and out-of-the class tasks.

“In the 5th grade, I use an eclectic-approach including dictation, questions and answers, and TPR in line with needs analysis. In the 8th grade, I teach the grammar in detail, translate the items or teach vocabulary as specified in the centralized exam in order for the students to do mechanical practices and rote-learning.... For example, I want the students to prepare a visual dictionary rather than rote-learning. I make an explanation that it may be time-consuming but

useful for them in the centralized exam. I do not tell them that they will learn English in this way. I just tell them that it will help them in test taking.” [LST-1]

In a similar vein, LST-V 2 notes that her way of teaching changes from eclectic methods to traditional ones, and she argues that this change stems from the coverage of the centralized exam. Regarding LST-2, LST-3, LST-V-1 and LST-V-2’s responses, it can be deduced that USE-EE brings vocabulary, grammar and reading comprehension into the spotlight; hence, communicative activities are less used in the 8th grade by comparison of the ones in the 5th grade; and test-taking strategies are practiced with the 8th graders.

“Through the 8th grades, there is a shift in language teaching from communicative and eclectic methods to traditional methods because vocabulary and grammar are important in the centralized exam. [LST-V-2]

“We use more [communicative] activities in the 5th grades. Especially in the 7th and 8th grades, the number of these activities decreases. It is because of the centralized exam.” [LST-V-1]

“There are some changes in the 8th grade. Our priority becomes the centralized exam. We focus on the exam vocabulary, item types, test-taking strategies in language teaching. On the other hand, I still use songs if they are related to the course content or I use word games in our revisions.” [LST-2]

“When I teach in the 5th and 6th grade, I feel happy because I [plan to] teach them practical English. However, in the 8th grade, I put this practical English [language] aside and teach them exam-oriented English language... In other words, the students, who will take the centralized exam, do rote-learning and practice reading comprehension rather than speaking and listening practices. This completely impedes our learning that we have progressed so far.” [LST-3]

LST 3 also makes a point that learning progress is also inhibited with the test-preparation practices. What’s more, PLST 2 calls attention to how language teaching differentiates in the 7th grades and onwards and explains in her response below that spoken production decreases in an educational setting where exam anxiety begins.

“We give more importance to speaking in primary education. We follow the same procedure in the 5th and 6th grades in lower secondary education but we have the centralized exam anxiety in the 7th grades. The students want to practice more language test items and do rote-learning.” [PLST-2]

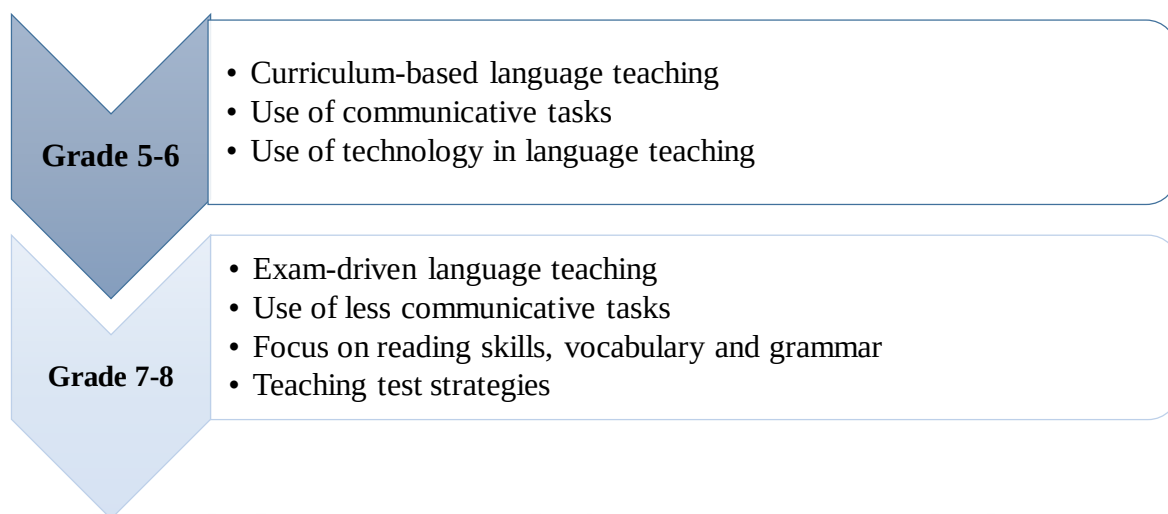


Figure 12. Language teaching shift in lower secondary education

Figure 12 describes how pillars of language teaching change through the 8th grades, at the end of which USE-EE is administered. It is concluded that teaching curriculum is narrowed into reading-based activities including grammar and vocabulary in tandem with USE-EE coverage. Exam-driven teaching overshadows use of communicative tasks and integration of technology into teaching.

Use of Supplementary Materials

Language teaching in the 8th grades evolves around exam-oriented instruction, strategy-based training and enhancement of reading skills. In line with these practices, EFL teachers employ different sources, especially test materials to get the students familiarized with different and well-developed language items in case similar items might be measured in USE-EE. To exemplify, LST 2 uses various supplementary materials, e.g. test books, to support their instructional practices in accordance with the changing item formats.

“Certainly, I use different supplementary books. Since it [testing system] changes completely [and] includes reading comprehension [not based on recalling], I prefer test books including well-structured items or I examine the language items in the centralized exams and teach students how to take these items.” [LST-2]

This response confirms that USE-EE motivates EFL teachers to bring different materials into classroom. Therefore, they prepare their students for how to deal with the centralized test items. Consequently, the students can perform well in the centralized exams.

Teacher Behaviours

In addition to its triggering force in the methodological shift, USE-EE influences language teachers' motivation, teaching coverage and preferences. In respect to motivating aspect of this centralized exam, PLST 3 expresses that she feels motivated to teach English language because of USE-EE. Test content and specifications enable EFL teachers to make sure whether all the units and objectives have been achieved or not.

"You put effort [into preparing the students for the exam] throughout the year. So, we expect the students to meet our expectations in the exam. Otherwise, it will be impossible [for us] to teach all the ten units with the same motivation. Even if there is only one item for each unit [in the centralized exam], we do a lot of practices related to this unit." [PLST-3]

Even though the centralized exam has a motivating touch on teaching behaviours, it may cause the anxiety among teachers in terms of preparing the students for this high-stakes test. LST 1 explicates that EFL teachers feel the sense of responsibility to teach effectively in the 8th grades. This responsibility results in unwillingness to teach in these exam-preparation grades because they feel under pressure as test outcomes will have high-stakes consequences. This affective dimension is exemplified in LST 3's response. He reveals his concerns and fear of covering all the points given in the curriculum by assuming that they might be tested in USE-EE.

"My colleague, EFL teacher, avoids teaching in the 8th grade. She thinks that teaching [English to the 8th graders] requires responsibility... She feels comfortable when she does not teach in the 8th grade." [LST-1]

"There is a fear of not covering all the units given in the curriculum. What happens if it [the unit or topic not covered in the lesson] is tested in the centralized examination? Due to this [feeling of] fear, we have concerns about covering all the curricular content." [LST-3]

Suffice it to say that USE-EE promotes teacher motivation but at the same time puts language teachers under pressure in terms of exam preparations. As this examination yields significant outcomes for all the parties, EFL teachers hesitate about taking such a responsibility of teaching in the 8th grades.

Exam-Driven Learning Expectations

The impact of USE-EE on language teaching is not only recognized in lower secondary education but also traced at upper secondary schools. UST-A-WE 2 expresses that the students in the 9th grades begin upper secondary education with their previous learning habits. They are accustomed to test-taking and doing mechanical activities; that is why they expect EFL teachers in the 9th grade to teach in a similar way. What is assessed becomes valuable for them to learn. This is regarded as the consequence of exam-oriented language learning in the 8th grades. UST-V-E 1 notices that the students want to give attention to the specific content that is going to be assessed and evaluated. It shows that exam-oriented teaching becomes the students' expectation in the 9th grade since they have undergone this kind of training in USE-EE preparation classes

“Students get astonished and nervous when they begin studying at upper secondary schools because they are used to mechanical exercises, multiple choice questions, tests to pass the centralized exam for transition to upper secondary education.” [UST-A-WE-2]

“In the 9th grade, at the beginning, they focus on the examination since they are not accustomed to the system [in upper secondary school]. They ask me whether I will assess the content in the examination and they need to put a mark there or not... They do not want to learn anything that is not covered in the exams” [UST-V-E-1]

Besides their expectations about exam-oriented language teaching, they ask EFL teachers to suggest test books that will support their language learning. UST-S-E 2 emphasizes that the influence of USE-EE continues and brings about the demand for use of test books. Similarly, UST-A-E 3 highlights the motivating aspect of scoring in her classes and again underlines their students' demand on using test books.

“[In the 9th grade], at the beginning of this year, a student asked me if there was a test book that I would recommend. I [explained] that there was no need for test book. The influence [of the centralized exam] is still going on.” [UST-S-E-2]

“In the 9th grade, in the first months, the students ask my recommendation for test books. When I give them homework, they ask me whether I score them or not”. [UST-A-E-3]

Putting together, the influence of the centralized exam for transition to upper secondary education can be noticed at the beginning levels of upper secondary education. The 9th graders have a tendency to attend only the activities that are scored and therefore assessed by teachers. They expect to be trained through languages tests including multiple choice

questions. This proves that USE-EE's influence may be long-lasting and exerted even after the administration.

Washback Effects of USE-EE on EFL Assessment Practices

Methodological Shift in Language Assessment

As well as its agentic role in bringing about instructional shift in what and how to teach English language, USE-EE causes some changes in language assessment methodology in upper secondary education. LST-V 2 teaches and assesses both the 5th and 8th graders, and explains below how different assessment methods she adopts in these grades:

“In the 5th grade, we begin with modern [alternative] methods like performance-based assessment [and] observation. We give more weight to formative assessment and encourage students to do projects. These activities are completely enjoyable and learner-centered. However, the assessment is getting traditional through the 8th grades. The rationale behind this is there is a centralized examination and I want to prepare them for this examination.” [LST-V-2]

More specifically, LST-V 2 conducts formative assessment. She administers performance-based assessment, observation, and projects as alternatives in assessment. Nevertheless, language assessment becomes traditional in the 8th grades due to upcoming USE-EE. LST 1 also validates LST-V 2's views and maps out her assessment practices by highlighting this methodological shift. That is, she utilizes observation, portfolio and performance-based assessment in the 5th grade. The 5th graders practice age-appropriate self-assessment tasks in addition to communicative tasks. However, she employs less communicative tasks in the 8th grade. The 8th graders get to know mechanical item types in the written exams.

“In the 8th grade, it is like a traditional examination. [My exam] includes matching, true/false, reading comprehension – questions and answers. I have to begin with these practices.... In the 5th grade, I do not administer a formal examination at the beginning. I observe them for 5th diagnostic purposes They have a portfolio for each unit. As this portfolio is the record of the learning process, it is a performance-based method in which I can see how the learners have progressed. Obviously, my first grade is about in-class participation in the activities. I consider how they perform in the following activities [given in the final sentence]. Second grade is about [the use of] these portfolios, and third grade is related to the general manner. The concept maps, portfolios, and worksheets are the [activities] used in the 5th grade. I prepare a self-assessment

form at the beginning of the academic year. For example, in the 5th grade, they do it by coloring the shape like the shape of an apple. [The shape is divided into small pieces according to the course objective and the learners color these objectives that they have attained.] [LST 1]

As for the alternatives in language assessment, LST-V 1 and LST-V 2 employ assignments and role-play tasks to assess the 5th graders. LST-V 2 also takes an advantage of anxiety-free assessment settings in the 5th grades and does not allow the traditional assessment to mandate the learning process.

“I have an assessment journal in the 5th grade and the students recognize it. I use it in order to assess their homework [assigning a plus or a minus] and give them feedback through it.” [LST-V-1]

“They do role-plays in these grades. It is important to have the learners enjoy the lesson and develop positive feelings for the English language. So, you cannot expose them to the written exams or mock tests directly. It becomes easier and more enjoyable. As there is no [large-scale or centralized] exam anxiety in the 5th grade, it is more comfortable. [LST-V 2]

On the other hand, PLST 2 draws attention to their attempts in getting the 8th graders familiarized with the centralized examination and puts an accent on the increment of exam anxiety through the 8th grades.

“We try to get accustomed to the [centralized exam] in the 7th and 8th grades. The students have exam anxiety. Along with gap-filling, matching and reading comprehension, we employ [language] tests.” [PLST-2]

“There is a variety of the activities [items] in the written exams [in the 8th grades] such as matching, gap-filling, and comprehension questions. Besides, there is a section with 25-30 multiple choice items. I try to prepare at least 6-7 different types of questions both to increase content validity and in order for the learners to practice [large-scale centralized exam].” [LST 2]

PLST 2 particularly uses gap-filling, matching, test items and reading comprehension questions in the 8th grade written exams. Likewise, LST 2 prefers similar item types in assessing the 8th graders. Plus, she puts effort to select different item types that can prepare the students for the high-stakes tests.

Aligning Language Assessment with USE-EE

In consideration of LST-V 1 and LST-V 2 responses, it can be understood that USE-EE is taken as a basis to shape in-class assessment practices. In other words, how influential the test specifications of USE-EE are in designing language assessment practices is seen in the following responses. LST-V 2 pays attention to USE-EE test specifications to measure the relevant items in her low-stakes assessment procedure. Put it another way, she develops the formal written exams in compliance with the centralized exam. LST-V 1's test-oriented approach derives from her teaching practices. Because of teaching English language within exam-oriented frameworks, she aims to get these outcomes in language assessment. It means that exam-oriented language teaching leads to exam-driven language assessment.

"We prepare our [formal] exams in accordance with USE-EE. All the exams to be conducted in the 8th grade include test items. We just consider what USE-EE measures. We try to assess and evaluate [language learning] on the basis of USE-EE." [LST-V-2]

"We adopt a test-oriented perspective in written exams. We expect to see how successful the students are in test taking or whether they understand the reading passages because we teach in this way." [LST-V-1]

In regard to the assessment of the language skills, PLST 2 touches on the pivotal importance of reading comprehension in the written exams. She attempts to align the written exams with the centralized test. On the other hand, PLST 1 pinpoints some ideas about the integration of four language skills into the centralized exam and asserts that they would test four skills if they were placed into USE-EE. This indicates that the inclusion of four skills in the centralized exam will promote and facilitate testing all the skills in lower secondary education.

"Reading comprehension comes into prominence inevitably in the 8th grades...I really check the items in my [written] exams whether they are appropriate to the level of the [centralized] exam or not." [PLST-2]

“Suppose that the writing skill is included in the exam or the students are required to produce a piece of writing or there are listening tests, we would obviously study [practice] those skills.”
[PLST-1]

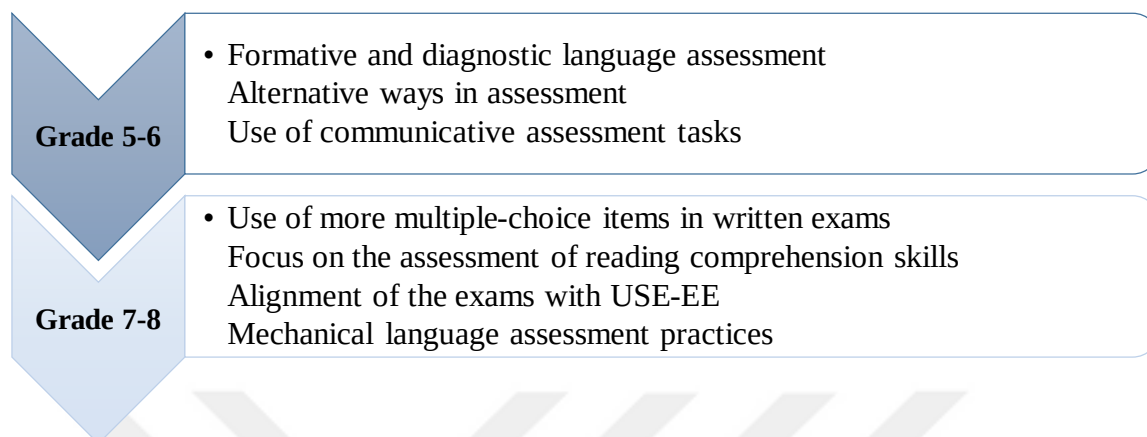


Figure 13. Radical shift between the grades in lower secondary education in terms of English language testing

Figure 13 illustrates how language assessment changes gradually in lower secondary education. In broad terms, assessment purpose is formative and diagnostic at the beginning levels. A large number of alternatives are utilized in language assessment such as portfolio, projects, and observation. In-class communicative tasks are widely used in assessing the students in these beginning grades. On the other side, mechanical exam items are commonly used in the 8th grades such as gap-filling, matching, and true/false statements. Some teachers make an endeavor to link their formal exams to USE-EE test specifications. Hence, they principally employ multiple-choice items and attach importance to testing the reading comprehension skills and strategies.

Washback Effects of USE-EE on Policy-Making

Language policies are specified and come into effect on macro-level and micro-level scales. Considering lower secondary schools as a micro-level dimension, USE-EE influences institutional policies that are introduced to improve language test performance. PLST-2 claims that the 7th grade is the cornerstone of language teaching in lower secondary education; from this grade, students become exam-focused and act accordingly as required in USE-EE. It can be deduced from this response that English language has a pivotal role in transition to upper secondary education, and therefore EFL teachers aim to raise this

awareness among the 7th graders:

“From the 7th grades and onwards, the students are aware of the fact that they will take a [centralized] exam in the following years, and they will be responsible for English language; for this reason, English language becomes a lesson that paves a way for getting education at a good upper secondary school. Frankly, we attempt to build this awareness and want the students to act in accordance with this reality.” [PLST-2]

Besides promoting exam awareness, institutions adopt a couching system for high-stakes exam preparations. As LST 1 explicates, the student number is divided by the number of lower secondary teachers. The main purpose here is to guide the students to gain study skills and prepare them comfortable and optimal learning settings. Additionally, couch teachers follow the students’ daily activities closely in order to provide them with a regular basis for exam preparations.

“In our institution, we have a couching system. It is very common in MoNE. At the beginning of each academic year, we take decisions about the implementation of couching system. To be more specific, [let’s say] there are 60 students in the 8th grade. It is impossible for us to follow each student’s learning process [mock examinations]. So, all the teachers [become responsible for] a small number of students. We deal with the students’ reading, eating, sleeping habits and routines. We ask them about how they spend their whole day, how many tests they take daily, what they study. When they gain these habits, we deal with their mock test scores.” [LST-1]

PLST 2 gives another example that is in relationship with exam couching system. In her school, mid-term exams are conducted to check the students’ progress. When there is a decrease in the scores of learning performance, the examiners organize family visits to in order to discuss the reasons behind this decrease. This finding points out the collaboration between teaching staff and parents. It can be concluded from PLST 2’s response that these visits help the students improve their performance in the practice exams.

“For example, when we conducted a mid-term assessment last year, we observed that there was a decrease in the 8th graders’ scores. Our administrator asked us to organize family visits. We went to the students’ houses and talked about how the results could be improved. After such a monitoring process, the students performed better.” [PLST-2]

Teachers and administrators develop these exam-driven policies mostly in their academic meetings. LST 3 shares his views on the coverage of these meetings. Accordingly, there are two main agendas on which they make decisions: instructional practices and the centralized

exam. Awarding is one of the policies developed with respect to the agenda of the centralized examination. As it is obvious in LST 3's response below, learners who get higher scores in USE-EE are awarded by the school administration.

"In our meeting, we focus on two important topics: Education/teaching and USE-EE that is important for the students, parents and schools. Our school is a successful institution in scholarship exams and USE-EE. So, our principal gives importance to this [examination] process. When the students succeed in the exams, they get awarded. We focus on how to develop our performance in relation to this examination." [LST-3]

PLST 2 calls attention to the decisions and strategies that are shaped in their meetings. Teaching staff and administrators evaluate the current status about the attainment of learning objectives and the results of USE-EE. Based on these evaluations, they develop some testing policies for the following years like the administration of process-oriented assessment and mock examination. By the same token, LST 1 emphasizes that they take testing-related decisions in their meetings in order to make preparations for USE-EE. Mock examinations and supplementary courses are the focal points on which their decisions are grounded. LST-V 2 also acknowledges that they make exam-preparation in the supplementary courses run by her institutions.

"We make evaluations on whether [language] objectives have been achieved or not; how many correct answers the students have got in each subtest of the centralized exam. We conduct many mock exams, process-oriented assessment and give importance to monitoring practices" [PLST-2]

"In our meetings [held] at the beginning of the academic year, we discuss what we [can] do [for the 8th graders]. We plan to run supplementary courses and conduct mock exams. In these courses, we completely do practices in line with the centralized exam." [LST-1]

"We have supplementary courses. The aim here is to get the students to prepare for USE-EE. These courses are very important; therefore, I am very interested in them." [LST-V-2]

Field Expert 6's response supports the findings about supplementary courses run at lower secondary schools in line with exam-oriented policies. He also puts emphasis on MoNE's own policies for the centralized exams. For example, Board of Education has a measurement and evaluation center in which they develop practice tests including the item types that are similar to USE-EE.

“Even though there is a systematic curriculum in MoNE, this exam brings about out-of-curriculum practices like supplementary courses... MoNE has testing offices and develops the items with good quality.” [Field Expert 6]

Along with these learner-centered and teaching- and testing-based policies, some institutions pursue teacher-related policies like accountability and designation. More particularly, PLST 3’s response indicates that school administration makes teachers accountable for USE-EE results and expect them to analyze and evaluate the sources of the failure if there is any. Furthermore, language teachers who have expertise in exam preparation are designated to the 8th grades in order for the students to get well-structured exam practices as explained in PST 1’s response.

“We certainly discuss it in our meetings. ... If there is a failure, teachers are judged about the reasons for this failure. We make evaluations in this way.” [PLST-3]

“Considering the results of mock exams or scholarship exams, the administrators designate language teachers to the grades [Teachers who get higher achievement in those exams are appointed to the 8th grades]. I have experienced it in lower secondary education.” [PST-1]

Simply put, USE-EE is influential in shaping lower secondary institutions’ policies related to language teaching, testing and teachers. Figure 14 displays these policies from a holistic point of view.

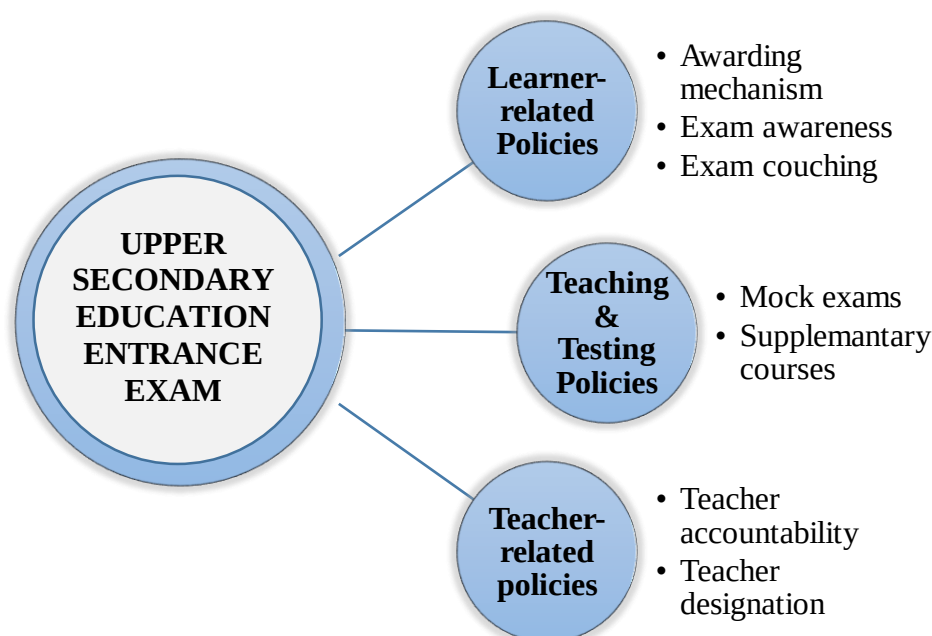


Figure 14. Institutional policies that are developed for USE-EE preparations

Washback on Language Teaching

- Methodological shift in language teaching
- Use of supplementary materials
- Teacher behaviours
- Exam-driven learning expectations

Washback on Language Testing

- Methodological shift in low-stakes language assessment
- Aligning language assessment with USE-EE

Washback on Policy-Making

- Teacher-related policies
- Learner-related policies
- Teaching and testing policies

Figure 15. Washback effects of USE-EE on language teaching, testing and policy-making

Figure 15 above gives an overall framework about how washback governs language teaching, testing and policy-making in lower secondary education.

Triggering Factors for Washback Effects of USE-EE

In the previous section, the washback effects of USE-EE were examined in detail with reference to language teaching, learning, testing, teachers and learners. This section gives an in-depth overview of social, political, economic and educational triggers of washback and their relations with realization of washback effects. There are six parts in this section: stakeholders' expectations and perceptions, socio-economic dynamics, institutional dynamics, low-stakes assessment policies, teaching duration and teacher-related factors. All the triggers have multi-dimensional relationships with different washback types as displayed in Figure 17. Obviously, high-stakes policies are the core and macro-level trigger that influences the direction, specificity and intensity of some second-level macro triggers and micro-triggers. Additionally, some triggers are also facilitated by other influential factors that can be regarded as underlying forces.

Stakeholders' Perceptions and Expectations

Stakeholders including school administrators, colleagues, parents and students have different perceptions about language teaching and expectations related to USE-EE preparations. These perceptions and expectations mediate how washback operates in lower secondary education. For example, in LST 1's school, the school principal does not take into account how successful the students are at productive skills; instead, he/she focuses on high-stakes test results and asks LST 1 not to use target language in teaching. As LST 1 discusses below, even if it is not suitable for her teaching beliefs, she is expected to carry out exam-oriented teaching. LST-V 2 mentions a similar point of view in her school community. The stakeholders expect her to support learning with exam coaching without considering the students' skills-based performance.

"When I began teaching in the 8th grades, the school principal asked me whether I used English language [target language] in teaching and I explained that I preferred using English. He recommended me not to use English... As it is my exam-preparation course, the stakeholders think that these students do not need to speak English. This is totally against the method I adopted." [LST-1]

"They expect us to motivate the students for USE-EE, do exam coaching for USE-EE, use language tests. They do not question whether these students are good at speaking or reading." [LST-V-2]

The colleagues from different subjects have also test-related expectations from language teaching. They expect EFL teachers to prepare the students to have higher scores in USE-EE because the students' good performance in English language sub-test will influence transition to upper secondary schools; thereby these students will be successful in other subjects, too. It is basically called inter-dependency. How interdependency affects teaching washback is elaborated with LST 1 and LST 2's responses below.

"One of my colleagues in the field of social sciences states that the student should not have four wrong answers in English language test because he/she [plans to study] at science upper secondary schools. That is, the success in one subject will affect the others." [LST-1]

"All colleagues believe that the students should be successful in all the subjects to pass the exam." [LST-2]

Like school administrators and colleagues, students come up with different expectations from their teachers in relation to exam preparations. Especially in the 8th grades, the students consider communicative tasks waste of time and unnecessary. Instead, they want to practice more tests as underlined in LST 1's response. LST-V 2 also calls attention to students' conceptions on learning and teaching. According to their perceptions, language teaching should be exam-oriented. What's more, in PLST 2's classes, the students expect PLST 2 to teach all the units that have a possibility to be tested in USE-EE, run more supplementary courses and do test practices even in testing.

"For example, we played games by using the smartboard but one of the students asked me whether we could take language tests. It is their expectations." [LST-1]

"The students think that teaching refers to taking tests and exams and learning grammar. As they are used to this process, they expect you to do so." [LST-V-2]

"For the students who will take USE-EE, the centralized exam is very important. They want to cover all the topics and have more supplementary courses. They expect us to use tests [in the written exams]." [PLST-2]

Parents' expectations and perceptions are similar to the students' expectations. Even if their kids are the 5th graders, parents ask teachers about how they prepare their students for the centralized exams as indicated in LST 3's response. Besides that, LST-V 1 stresses the parents' demand on recommendation of exam-preparation materials because they want their children to give a start to exam preparations in advance.

"In our parental meetings in the 5th grade, parents asked if we had any specific methods for USE-EE." [LST-3]

"It is related to the parents' points of view. If the parents expect their children to have high scores in USE-EE and study at a good upper secondary school, they visit [the schools] and tell me that they begin getting prepared for the exam. They ask my opinions for test materials." [LST-V-1]

As well as these expectations, the parents' perceptions on language teaching is exam-focused. They show resistance to use of communicative tasks; they believe that their children do not need to practice these tasks, and these tasks will not be influential in promoting exam scores as depicted in LST 1's response.

“For example, I sent a list of 50 songs to the parents and requested them to play these songs at background when their children play the games. One of the parents asked me why I sent them and told me that they were unnecessary... [Another example] I was teaching [the modal of] “can” and asked the student to color the activities that they could do... but one of the parents sent me a message in which she said that they were fed up with these activities and questioned the influence of these activities on learning English language.” [LST-1]

Field Expert 4 verifies that parents consider communicative tasks useless because language items are not presented in this way in the centralized exams. Therefore, they suppose that these tasks do not contribute to their children’s exam preparations. Field Expert 3 argues how these perceptions are shaped in society. He claims that the society’s primary concern is to assess the children. They do not mind what the students have learnt but how they have performed in the exams.

“The parents do not think games and songs are effective in teaching and they make contribution to the exams. They are partially right because the [centralized] exams do not measure the objectives that can be attained through these games and songs.” [Field Expert 4]

“Because we do not give very much attention to learning we are giving very much attention to assessing, we do not have a concern whether students are going to be constructive students, are students going to be critical thinkers, or are students going to be explorer, this is not what we are interested. What we are interested in very much like how we are going to assess the kids?”

[Field Expert 3]

Field Expert 6 explains why the society’s primary concern is assessment. He emphasizes testing purpose of high-stakes examination. These exams are administered to select the students for upper secondary or university education. When the students are selected at the end of these exams, they are viewed as successful language users no matter how successful they are in productive skills. He also gathers attention to stakeholders’ knowledge of language teaching. Another reason why the stakeholders adopt exam-oriented teaching might be their beliefs that language should be taught traditionally. Only if they are trained about how effectively foreign languages can be taught and what practices should be done, their perceptions may start being varied.

“It is because our society is exam-oriented. In society, we tend to select the individuals through the exams. Secondly, the parents, administrators or the society has insufficient knowledge about language learning, i.e., how languages are learnt, which activities are used and which skills are

taught. In our society, success is understood from the exam results. [For example] A student who has 100 points in the exam is accepted as a successful one even if he/she cannot speak or write [in target language].” [Field Expert 6]

All things considered, it is understood that school administrators, colleagues, parents and students expect EFL teachers to do exam-oriented practices in class and aim to increase the students’ test performance. Especially parents and students have some exam-focused beliefs about language teaching and testing. Large-scale and high-stakes testing system shapes these perceptions and expectations. The relationship between washback and the stakeholders’ expectations and perceptions is demonstrated in Figure 16 below:

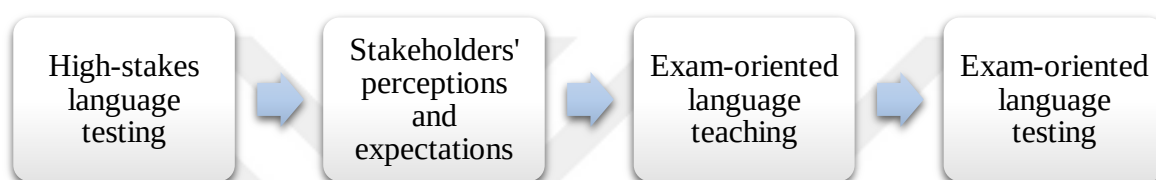


Figure 16. The relationship between exam-oriented language teaching and influential factors that trigger exam-oriented language teaching

Socio-Economic Dynamics

Another main trigger that exerts the washback of USE-EE on EFL teaching and testing dynamics is socio-economic factors. LST 1 puts emphasis on the financial force behind centralized exams. Parents believe that their children will be able to get a job when they pass the centralized exams [USE-EE and university entrance exam]. From their perspectives, getting a job corresponds to making money and affording their life. Thus, parents make a connection between the exams and economic wealth.

“Parents suppose that a good job means a good income. Our school is located in a suburban district. In the 8th grade, parents are not curious about what their children have learnt; instead, they view their children’s exam success as the life achievement. This [exam success] is considered a way for making money. The only purpose is to earn money and afford the life.”

[LST-1]

Considering socio-economic consequences of high-stakes testing, Field Expert 4 supports LST 1’s point of view and states that education is viewed as the best way to enhance the students’ social and economic status. That is why the parents expect their children to pass

the exams. In doing so, they could change their socio-economic status. In addition to getting a job, earning money and changing socio-economic status, the centralized exams play an essential role in getting the social acceptance. To take an example, when people become successful in the high-stakes exams and then have a prestigious profession, they can get approval and acceptance from the society according to LST 1's following response.

"Parents think that education is a tool to change the social status. The parents that have relatively lower socio-economic background believe that education is the best way to promote their children's socio-economic level. From this point of view, they expect their children to pass the exam." [Field Expert 4]

"In the society, people introduce their profession before introducing themselves because [these] properties determine their position in the society and enable them to get approval... Besides earning money, this [centralized] examination will meet their [other] needs. They will be successful in the university exam, then in life. What happens later? They will get accepted, approved and respected by the society" [LST-1]

In a nutshell, USE-EE's influence is also driven by socio-economic factors. Examination results pave the way for getting professional education, earning money and changing socio-economic status. Furthermore, it helps the students get social acceptance, approval and respect. For this reason, it gains importance in the society.

Institutional Dynamics

Lower secondary schools gain prestige on the basis of the USE-EE scores. Namely, students' good performance on test positively affect institutional dynamism as explained in PLST 2's response. At this point, LST-V 1 and LST-V 2 discuss that school prestige is influential in student recruitment. If a lower secondary school is prestigious, parents tend to get their students to register for these schools.

"The school gets prestige as a result of this examination. It affects the school dynamism and performance remarkably." [PLST-2]

"In consequence, the school becomes successful. Parents want their [children] who are the graduates of primary education register for this school." [LST-V-1]

"When [the 8th graders] pass the exam, our school gets prestige and the students with high level of academic achievement will register for our school." [LST-V-2]

Another trigger, at the same time washback itself, is the institutional comparison conducted at the end of the examination. The schools are compared at district and provincial levels according to the high-stakes test results. LST 1 shares her experience in relation to such comparisons by highlighting the institution's purpose to catch up with the average scores. Also, she draws attention to awarding system for the successful schools and teachers. It is expected to give a round of applause to EFL teachers whose school has the highest average score. Likewise, at the school where LST 3 works, the school principal attaches importance to the student's average score and its comparison with other schools. Field Expert 6 mentions his experience during practicum where he observed that the school administrators viewed USE-EE results as the criterion for success and compare their school performance with others considering these results.

"The scores are compared at the school, district, and provincial levels. We put effort into getting the nationwide average score.... If the average score in your school is below the one in the district and province, your school becomes unsuccessful. Also, in the meetings held in the district, [EFL] teachers, whose schools become successful, get applause from others" [LST-1]

"We compare the results at district and province levels. Our principal expects us to discuss the results and report our average scores." [LST-3]

"I used to visit practicum schools. In general, [school] principals used to claim that their school was one of the best schools in the city [they were proud of their success in USE-EE]. They regard this exam as a criterion to compare themselves with others." [Field Expert 6]

EFL teachers and Field Expert 6's responses prove that lower secondary schools are compared on the basis of USE-EE takers' scores and gain prestige if the results are high as expected. As a consequence, these prestigious schools become the institutions for which the parents prefer getting their students to register. That is, USE-EE enables the schools to select their own students.

Low-Stakes Assessment Policies

As elucidated in the argumentations in the previous chapter titled *"Methodological Shifts"* in both teaching and testing, USE-EE's exam coverage is one of the reasons why four language skills are not completely taught and assessed especially in the 8th grade. Another reason is attributed to the principles and requirements of the examination regulation as PLST

3, LST 1 and LST 2 summarize. The current examination regulation requires EFL teachers to administer a number of written exams officially. Although the ELT curriculum proposes the assessment of four skills in addition to formal examinations, teachers shape the assessment procedures in compliance with the examination regulation:

“We conduct written exams as many as the regulation requires us. We complete the assessment and evaluation procedure with these formal exams.” [PLST-3]

“We do not measure speaking and listening skills probably because the regulation does not require us to measure these skills... School administration examines if we conduct written exams in accordance with the [examination] regulation and we have got written exams and answer keys.” [LST-1]

“They expect us to develop the exams [including] different difficulty levels in accordance with the examination regulation and use various methods and item types.” [LST-2]

The fact that EFL teachers are required to follow the official regulation principles in low-stakes language assessment is also expected by the school administrators. LST 1 and LST 2 specify the school administrators’ approach towards language assessment. Evidently, there is no expectation of assessing four language skills because it is not included in the examination regulation. In this regard, Field Expert 6 emphasizes the necessity for policy development to assess four skills. If skills-based assessment becomes an official regulation, EFL teachers are required to implement it in language classes.

“How many teachers that assess four skills have you seen so far? We need policy development at this point. MoNE should require language teachers to develop the exams covering four skills. They can be also prepared by group teachers. It will be a tradition later.” [Field Expert 6]

Considering foreign language related policies in Turkey, Field Expert 1 indicates the official paper published by MoNE in 1983 and revised in 2006. As discussed below, how to teach language skills is explicated in this official paper. Field Expert 1 questions why EFL teachers do not measure the achievement of language objectives as given in this policy. This view points out a gap between teaching and examination regulations. Even so, he recommends his student teachers to use this policy as a guide in English language teaching and testing.

“MoNE has [published] foreign language teaching aims [1983;2006]. That is the official paper. [It says] you should teach four skills. How are you going to decide whether your students gain these skills or not? By assessment. So, I always tell the teachers that this is your starting point. This is the aim.” [Field Expert 1]

From an overall perspective, the rationale behind not assessing four skills is the lack of an official regulation and policy in low-stakes language testing. Field Experts believe that skills-based language assessment will be put into practice by the time it is placed into the examination regulation.

Teaching Duration

Exam-driven language assessment is an obvious washback effect of USE-EE. Time allocated for English language teaching is another dimension that demotivates EFL teachers to carry out communicative language assessment tasks. LST-V 1 and LST-V 2 especially note that alternative and communicative assessment requires the sufficient amount of time in terms of administration, and weekly lesson duration is only enough to conduct traditional methods.

“We do not have [enough] time to give performance-based assignments or do projects. That is why I adopt more traditional perspectives through the 8th grade.” [LST-V-2]

“It changes from school to school. In our school, there are weekly ten lesson hours in the 5th grade; on the other hand, it is reduced to three hours in the 6th grades. Thus, whereas we employ tasks in the 5th grade, we focus on teaching vocabulary and grammar and do not use so many tasks [in the other grades].” [LST-V-1]

It means that EFL teachers need more time for four-skills assessment, alternative and communicative assessment; and lack of time causes EFL teachers to follow more traditional techniques and methods.

Teacher Professional Identity and Assessment Literacy

In primary education, teacher-related factors are indicated as the elements mediating the washback. However, in lower secondary education that holds a high-stakes test: USE-EE, teacher-related factors function as the triggering factors that govern the washback effects. In particular, one dimension of these factors is the assessment literacy, which refers to EFL teachers' knowledge and skills of assessment that are influential in designing their assessment practices. For example, LST 3 echoes his assessment manners and states that he aims to assess different language skills. On the other hand, PLST 1 and PLST 3 comment on their practices and explain that they do not assess some certain skills.

“I wish to prepare reading passages, listening parts, and writing guides for my [formal] exams. Since I was appointed to MoNE, I have not focused on the grammar.” [LST-3]

“I have never assessed the listening skill in my formal exams. Some teachers assess it but I do not.” [PLST-1]

“Some teachers assess listening skills in the exams but, to be honest, I do not assess this skill [in the formal exams].” [PLST 3]

The responses show that EFL teachers differ in their assessment practices, and some teachers do not prefer assessing some language skills like listening. Even though their assessment preferences may vary, LST 3 mentions the identical formats of the written exams that are developed by EFL teachers and shared on an internet-based website. He claims that most language teachers use others’ written exams without changing any items. It puts evidence on the fact that EFL teachers need to improve their testing knowledge and skills in order to develop various item types for their own assessment procedures.

“More specially, in the 8th grade, there is a website on the internet for sharing written exams. Can you imagine all exams are similar to one another? Item types are the same. So far, I have tried to use different practices. For example, I recorded my voice and prepared a listening material for my exams.” [LST 3]

Field Expert 8 makes a point here about what shapes EFL teachers’ assessment practices in light of her interaction with pre-service teachers. She asked her students about how testing should be done and received the answers in consistence with testing practices to which they were exposed when they were language learners. She finds out that their educational background determines how language assessment should be conducted.

“When I ask my students again, what makes you drive whether testing foreign language should be done in x or y way. What I realized from their answers was a combination of a number of different factors. It looks as if one of the influential factors was their own experiences as language learners. What type of testing worked for them or what type of testing motivated them while they were language learners, it looks as if they think, they perpetuate their beliefs? Their own experiences as language learners.” [Field Expert 8]

Regarding the power of language teachers’ educational background, Field Expert 5 suggests assessing the student teachers’ performance in tandem with techniques and methods with which they are instructed to follow in pre-service education. She stresses the experiential aspect of alternative assessment and underscores that pre-service education should present

them with opportunities to gain an understanding of alternative assessment and practice it at tertiary level. LST 1's response below supports Field Expert 5's perspective and reveals that she could not have any practical experiences before in-service teaching.

"Pre-service teachers should be familiar with alternative assessments in their own courses... I absolutely show some examples of alternative assessments... I believe that field-specific examples are very useful; without those examples, alternative assessment is not [fully understood] by them. Without experiencing it as a learner, they cannot put it into practice."

[Field Expert 5]

"We got [assessment] [training at university. Yes, our teacher trainer taught us how to conduct language assessment but we internalized it after we began working." [LST-1]

Besides assessment literacy, teachers' professional identity also has a triggering effect on how washback on teaching is handled. Some teachers may be knowledgeable about how to teach and test language skills effectively. However, their professional identity directs them not to push themselves and distort their comfort zone of teaching as in LST 1's expression. In addition, some teachers do not resist against the system as LST-V 2 exemplifies below. Instead, they allow the effects of USE-EE to mandate language teaching and testing.

"It depends on teachers' own personal perceptions or identity. I always take the easy way out. I comply with the stakeholders' expectations." [LST-V-2]

"There are some teachers who will not change their practices even though they get in-service training because they follow the practices in which they feel comfortable." [LST-1]

"If there are some challenges, if there are some demanding tasks for the teachers so they do not want to do that, because they feel that they are safe and their teaching is enough for their students. So, you cannot change these perceptions of language teachers [and] other subject-field teachers so the thing we should do is to prepare some guidelines of testing, and we should offer them opportunities to improve themselves about testing, otherwise we cannot change the testing system in our schools." [Field Expert 2]

Another identity-related issue is that some teachers may not be adaptive to the assessment tasks that require extra effort and energy. They will continue in a way that they feel safe as explained by Field Expert 2. In this regard, he recommends offering testing guidelines to EFL teachers in order to improve their language assessment practices. This response again promotes the importance of introducing newly-designed and updated testing policies into in-service education.

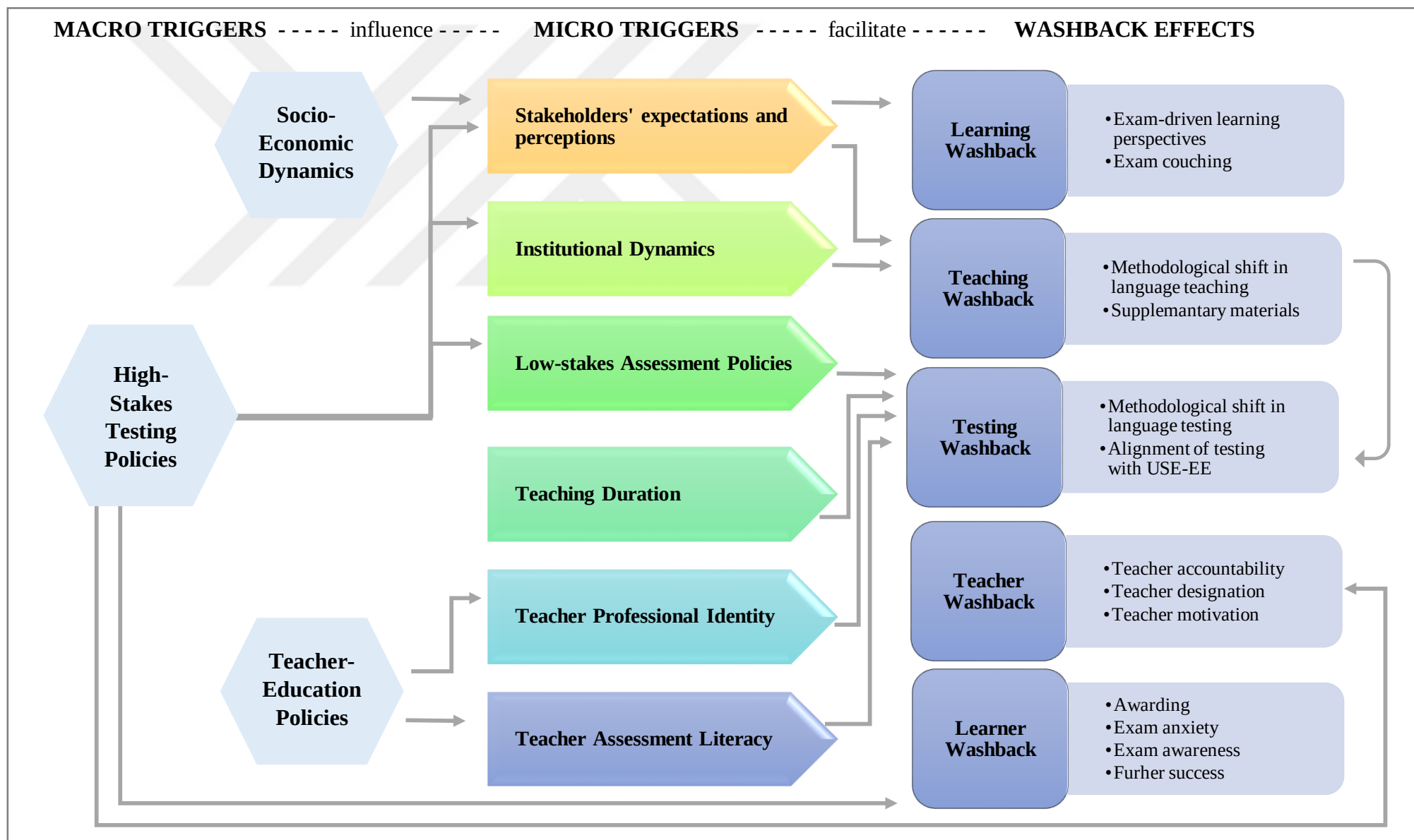


Figure 17. Operationalization of USE-EE washback effects triggered by macro- and micro-level dynamics in lower secondary education

Case 3: Upper Secondary Education

As the third case of this research, upper secondary education is analyzed thoroughly in terms of washback effects of low-stakes and high-stakes assessment and washback-triggering forces. There are two basic parts of this chapter: (1) washback effects low-stakes assessment and triggering forces and (2) washback effects of high-stakes assessment and triggering forces. The first part deals with teacher-based washback effects on language teaching, learners and teachers and puts emphasis on examination regulation and diploma scores as triggering factors. In the second part, both BCT and FLT are examined in regard to their impacts and consequences. Additionally, this part illustrates why the centralized exams exert washback effects in upper secondary education.

Washback Effects of Low-Stakes Assessment and Triggering Factors

Teacher-based assessment refers to classroom assessment practices that are principally developed, conducted, scored and reported by EFL teachers, and it is called “low-stakes assessment” since it is geared to micro dynamics of EFL teaching, and its impacts are interpreted in relatively micro settings. These practices exert some effects on language teaching dynamics. In particular, they promote learners’ spoken production, provide feedback on language learning and help revising teaching procedures, influence language learners’ habits and affective development, increase teacher collaboration but impede teacher autonomy. These washback effects are explicated in the first section. Following that, the second section presents the triggering factors that facilitate the emergence of these impacts in upper secondary education. These triggers are basically categorized as testing policies. While teachers’ responses are the main source of qualitative data for low-stakes assessment washback, both teacher and expert views are examined to find out the triggers.

Washback Effects of Teacher-Based Assessment

Enhancement of Spoken Production

With the introduction of performance-based assessment (skills-based assessment) into upper secondary education, EFL teachers administer skills-based examinations. UST-A-WE-1 talks about the influence of this speaking examination on the students’ speaking performance. She says that her students enjoy attending and completing assessment tasks. These tasks promote the students’ spoken production.

“I used to give them some topics and ask the students to choose one of them in the 9 grade. In the 10th grade, I asked them to introduce Turkey or one of the cities in the world and made a presentation about it. They felt happy. They enjoyed preparing presentation and made effort [to speak].” [UST-A-WE-1]

“Normally, we cannot enable 30-40 students to take turns in speaking but we do it during speaking assessment. Probably they become aware of what they can do. Let’s say 10 students attend in-class activities but other 20 students learn in a passive way. These [twenty] students may understand that they can speak or achieve something [during or at the end of assessment.].” [UST-A-E-1]

Contributing to the improvement of speaking skills, performance-based assessment helps the students test their learning hypotheses and come to know what they have achieved at the end of assessment as understood from UST-A-E 1’s response. UST-A-E 1 also pays attention to testing time and asserts that the students have more time to produce spoken production in assessment when compared to in-class practices. It is safe to say that teacher-based assessment enhances the students’ language skills, especially speaking skills. The students who have enough time to use English language in assessment procedures realize what they learnt or how they use the language.

Feedback on Language Learning

During or at the end of language assessment, students sharpen their understanding about what they have gained in learning. Sometimes, when they have low scores in assessment, they believe that they are not good language users. It shows that the students make judgments on their performance in line with the assessment results. These judgments may also affect their in-class participation.

“I have been working at this school since last year. When I met the students, I realized that the students participated in the lesson for the sake of being scored. When they did not attend the activities, I asked them the reasons why they did not take participation. They answered that they were not good at English lesson. I have deduced from this answer that they make judgments on whether their level of English [language] is ‘good’ or ‘bad’” [on the basis of exam scores].”

[UST-A-WE-2]

Assessment results are not only interpreted by language learners but also analyzed by teachers at grade- and school-levels. UST-A-E 2 states that they revise their teaching in

accordance with learning strengths and weaknesses that are determined on the basis of the assessment results.

Exam results are analyzed. We get the analyses and examine learning strengths and weaknesses. In our meetings of group teachers, we make decisions on doing revisions or various exercises.”

[UST-A-E-2]

“In the meeting held at the end of the academic year, we discuss the differences [in success] between the grades and talk about what we should do. [UST-A-E-1]

“Exam results are analyzed. According to these analyses, we make comparisons among the classes. [UST-S-E-1]

Similarly, UST-A-E 1 explains that EFL teachers argue the grade-based differences in language learning in light of the assessment results and develop some plans and strategies accordingly. UST-S-E 1 also confirms that grade-based comparisons are made in line with the analyses of exam results. It is seen that teacher-based assessment results serve as a basis for getting feedback on language learning and planning further steps to improve language teaching. Additionally, these results are used to make comparisons among the grades.

Affective Outcomes

Teacher-based assessment has some affective influences on language learners. For example, even though skills-based assessment is effective in improving spoken productions, it causes excitement among the students. UST-A-WE 1 mentions that their students feel excited during speaking examination; for that reason, she does not score their performance severely in order not to demotivate the students.

“MoNE requires us to assess speaking skills now. The students make effort [to speak] but get excited. I give them the topics... but they have tremendous excitement. I do not assign low scores to their performance in order not to reduce their motivation.” [UST-A-WE-1]

“Unfortunately, the system focuses on exam and score. Although I explain that English language is not a lesson, and it is beyond the lesson, we have to conduct the exams and score [the students]. We have to test what we teach. This causes anxiety [among the students]. Anxiety affects most of them.” [UST-A-E-1]

Another affective outcome of teacher-based assessment is the feeling of anxiety. UST-A-E 1 believes that English language is not a regular lesson; nonetheless, they have to assess what

is taught in the classroom. She thinks that this requirement of assessment causes anxiety among the students. To sum up, excitement and anxiety are two main affective aspects of teacher-based assessment washback on language learners.

Exam-Oriented Learning Habits

Students' habits of learning are also shaped by teacher-based assessment. Their understanding of learning becomes equivalent to being scored or assessed. In general terms, the students want to have exam-focused learning experiences. In this regard, UST-A-E 3 mentions that her students can only do their homework if it is taken into account in scoring. Drawing attention to common implementation of the exams, UST-A-WE 5 poses that teacher-based assessment, especially common exams, makes the students exam-oriented. Their parents perceive the common exam results as the standard of achievement. For that reason, the stakeholders focus on examination process more rather than learning process.

"When I give them homework, they ask me whether I will give [them] 'plus' or 'minus' or how I will score them. They are used to [this kind of assessment]. If it is not scored, the students do not do their homework." [UST-A-E-3]

"They [students] become exam-oriented. Parents accept the common exams as the criterion of the success." [UST-A-WE-5]

"The students tell me that they have difficulty in pronouncing the words, acting like those in the videos they watch or they want to develop speaking skills. It makes me happy and I see the importance they attach to language learning. However, through the exam weeks, they realize that the exams will be conducted and ask me what topics they should practice [to pass the exam]." [UST-V-WE-2]

Likewise, UST-V-WE 2 touches upon the same issue. In her school setting, the students are interested in developing language skills, and this manner pleases their EFL teacher. However, through examination time, the students become more interested in what is going to be assessed. Briefly, teacher-based assessment influences learning behaviours and perceptions and thereby facilitates exam-driven learning habits.

Teacher Collaboration

The implementation of common written examinations leads to teacher collaboration at the upper secondary schools. UST-A-WE 2 expresses that group teachers work in cooperation

to develop formal written exams. First, they take decisions about examinations and then prepare their own exams on the basis of workload. That is, each grade to be assessed is assigned to one EFL teacher, and they prepare the exams individually in advance. Then all teachers give feedback to one another about the exams.

“[EFL] teachers can conduct performance-based assessment by themselves and prepare different kinds of questions. On the other hand, written exams are specified as the common examinations according to group teachers’ decisions.... We share the grades to prepare the written exams. While Teacher A develops the exam for the 9th grade, Teacher B develops the exam for the 10th grade. After the development of the exams, all teachers give feedback to each other about all the exams.” [UST-A-WE-2]

Similar to UST-A-WE 2’s assessment settings, in UST-A-WE 3’s school, EFL teachers get feedback from each other about the exams they have developed. UST-A-WE 3 pays attention to their group meetings and their discussion about language assessment. Furthermore, UST-S-E 1 puts emphasis on EFL teachers’ interaction in the schools and explains that they share even the vocabulary they have taught with each other because they will develop a common written exam, and they might measure this vocabulary there.

“We talk about assessment and evaluation regularly [in the meetings of group teachers]. We discuss what we have done in the [written] exams and how they were resulted. ... We certainly share the exams with each other. If I prepare them, I ask for others’ opinions. When they develop them, they ask for my opinions [about the exams]. Otherwise, the students make objections [when they come across the questions related to the content that is not covered in the class].” [UST-A-E-3]

“We have a good communication [and] talk about even the words we have emphasized during the lesson because we may ask them in the written exams. All the exams are common in our school. Otherwise, it will cause problems because there is serious exam anxiety, and all the stakeholders interfere with this process.” [UST-S-E-1]

“We get in touch with each other and prepare rubrics about how to score [the performance] in performance-based examination. We share [the tasks of] the written exam preparations and then we review these exams. For example, we discuss how to score the reading comprehension.” [UST-A-E-2]

Along with the development of written exams, UST-A-E 2 accentuates their cooperation in how to score both written exams and performance-based outputs. To this aim, they prepare

rubrics together. Like other EFL teachers, they also develop common written exams together. In short, common examinations promote teacher collaboration by enriching teacher interaction, cooperation and feedback-giving. They work together in not only developing written exams but also scoring these exams.

Teacher Autonomy

As discussed above, common written exams affect positively and dynamically teacher collaboration in terms of interaction, feedback and cooperation. In essence, this examination was put into practice for different purposes, one of which is to reduce variance occurred in teachers' testing behaviors as given in UST-A-WE 4's response.

"We here conduct the common exams in order to minimize teacher-related factors in assessment." [UST-A-WE-4]

"In the past, every teacher was responsible for their own assessment. Now, we have to conduct it altogether [the common examination]. Is it good or bad? Actually, it was intended to [build a standard] for good purposes but you have to comply with the teachers who follow the curriculum behind. You have to ask moderate-level questions. You cannot ask the important points that you emphasize in the classroom." [UST-A-WE-1]

On the contrary, UST-A-WE 1 questions whether this implementation is useful or harmful. She clarifies that it is an intentional attempt made for beneficial purposes. On the other hand, it affects teacher autonomy negatively because all teachers have to comply with others' practices. For instance, when teaching content is partially covered in a class, common exams should be prepared in accordance with this content.

Triggering Factors for Teacher-Based Assessment Washback

High-Stakes and Low-Stakes Assessment Policies

Teacher-based assessment refers to both formal and performance-based assessments that are determined in the examination regulation as elucidated in UST-A-E 1. These exams should be conducted commonly by EFL teachers. Such a requirement given in the regulation facilitates teacher collaboration among the students.

“What we are required to do [in assessment and evaluation] is given in the regulation. It says that at least two formal exams need to be conducted. These exams are divided into written exams and performance-based exams.” [UST-A-E-1]

“Performance-based assessment became mandatory in MoNE regulation [for upper secondary education]. There are skill-based exams in theory.” [UST-A-WE-5]

In addition to the triggering influence on teacher collaboration, the exam regulation promotes the use of skills-based assessment as explained in UST-A-WE 5’s response, and this leads to improvement of speaking skills. As another washback effect, teacher-based assessment has some affective impacts on learners such as excitement and anxiety. In this respect, UST-S-E 2 argues that the influence of low-stakes assessment on university-entrance exam results causes exam anxiety among the students. In other words, at which university the students will get tertiary-level education depends on their diploma score to some extent. The students who have this awareness feel anxious about their exam results.

“You know that there is diploma score in our system. The higher the diploma score is, the better university you are placed in. Since they [students] have this awareness, it is very normal that the students have exam anxiety.” [UST-S-E-2]

“Classroom-based assessment/exam scores will influence university [entrance] examination scores, and parents are aware of [this influence]. That is why they care about the exams. Students and parents have more concerns on assessments because they will affect the centralized university examination.” [UST-A-WE-1]

UST-A-WE 1’s response shows consistency with abovementioned perspective. She claims that not only students but also parents are aware of the link between teacher-based exam results and diploma scores. For this reason, teacher-based examination affectively influences the stakeholders. This link also makes learners exam-oriented as UST-A-E 1 clarifies. It can be deduced from her response that testing policies drive the students to take part in the courses because the content taught can be assessed, and assessment results play a key role in shaping the diploma score and thus university entrance.

“They think that they should attend the course because of the exams. In their opinions, teachers will assess what is taught in these courses. I will get the scores from this assessment. These scores will define my achievement and influence my transition to university.” [UST-A-E-1]

To conclude, examination regulation and diploma score are two basic influential factors that trigger the low-stakes assessment washback on learning, learners and teachers. Figure 18

gives an overall framework of how washback effects of teacher-based assessment interplay with EFL teaching dynamics and what forces govern these effects. Evidently, testing policies are the underlying factors that formulate how low-stakes assessment practices are to be carried out. As seen from the Figure 18, these practices influence teachers, learners and language learning.

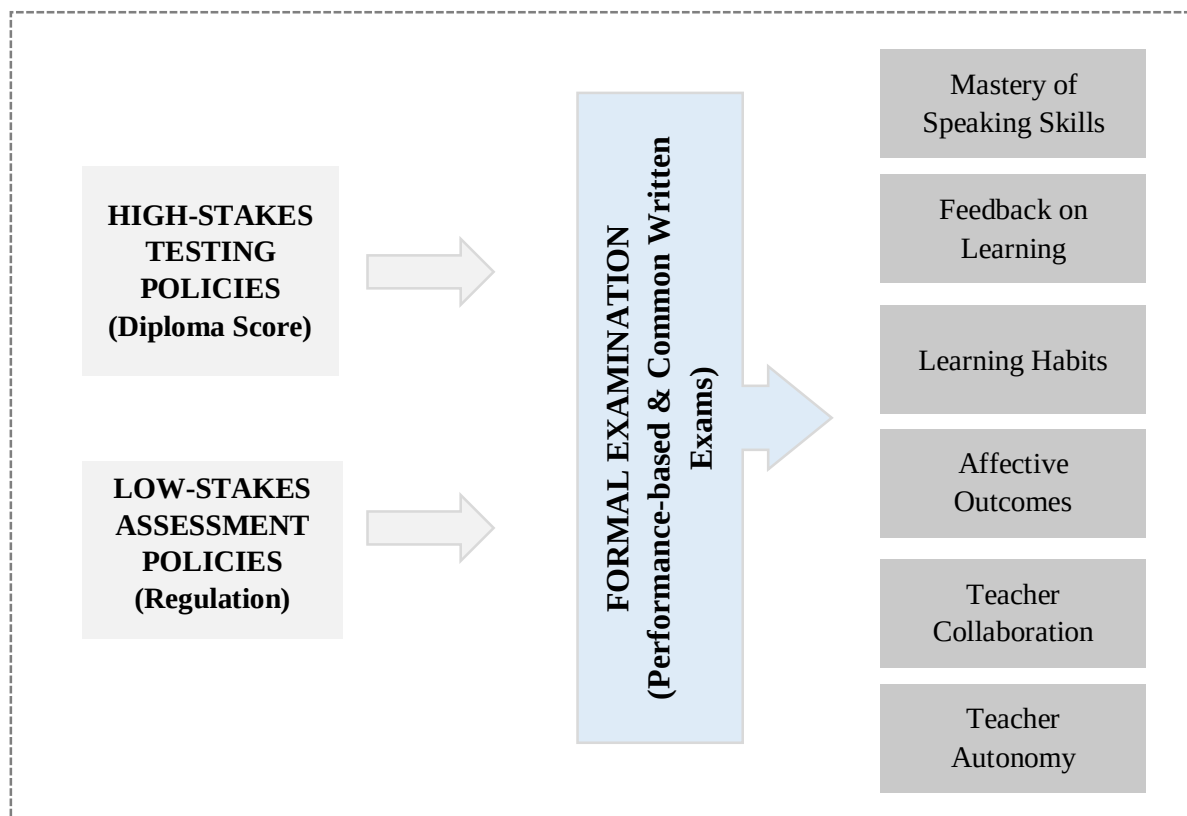


Figure 18. Washback effects in upper secondary education and main triggers that exert the effects of teacher-based formal examination

Washback Effects of High-Stakes Assessment and Triggering Factors

This section scrutinizes the washback effects and triggering elements of two main high-stakes tests: BCT and FLT and. Firstly, BCT effects are elucidated with respect to teacher and learners' motivation. Following that, the effects of FLT are discussed in relation to language teaching, testing and policy-making. In the third part, both macro and micro triggering forces are displayed under five themes: stakeholders, socio-economic and cultural dynamics, institutional dynamics, high-stakes test blueprints, and teacher-related factors.

Washback Effects of BCT on Language Learners and Teachers

Motivation Loss

At the end of the 12th grade, the students take BCT that is first administered test to measure basic skills for university-level education. In BCT, there is no section for English language; so, the non-English major students (the 11th and 12th graders) are not willing to learn English language. UST-A-WE 6 presents her observation that only the subjects measured in BCT become important for the students. Even though English language is essential in determining diploma score, it is not taken into account by these students.

“I do not think that the 11th graders (except for language classes) give importance to English language lessons. As far as I observe, the lessons measured [in BCT] are important for the students. Even though [English language lesson] it plays an essential role in diploma scores, [the 11th graders] ignore it because they have no motivation. They plan to learn it later [after university entrance examination]. [UST-A-WE-6]

Similarly, UST-S-E 1 calls attention to science upper secondary schools and claims that there is not motivation to learn English in the 12th grades because they have demands on free time to practice other subjects. Although EFL teachers make decisions on teaching English in these grades effectively, they cannot show resistance to the students’ demands. In relation to their demands, UST-A-E 3 mentions that they made a contract with the students about giving free time and teaching English language. UST-A-E 1 brings another perspective into the interpretations of the students’ demands with regard to learning priorities. She expresses that students do not have the need to learn English because their priority is to pass the exams of other subjects. Due to this priority, the students use test materials of other subjects in the classroom, and she finds no way to change their priority.

“The 12th graders think that our teachers give them freedom [to focus on university examination] in the lessons. It is the same at all science upper secondary schools. Each year, I decide on teaching effectively but in the 3rd or 4th week, I give up then.” [UST-S-E-1]

“We made an agreement at the beginning [of the term]. We taught them English for two lesson hours and then we gave them freedom to take tests [of other disciplines] for other two hours. [UST-A-E-3]

“The students have different concerns [in the 12th grades]. They claim that language learning is not their priority at the moment. They say that they do not want to learn English now because they will take an exam [BCT or FCT] and they are expected to perform in [other subjects]. They

bring test books [of other subjects] to the classroom. How can I change these students' priority?" [UST-A-E-1]

The exam coverage of BCT results in motivation loss among the 11th and 12th graders in non-major English courses. They attach importance to only the subjects that are measured in BCT and FCT because the students' priority is to pass these exams for transition to university education. As the students are not volunteer to learn English in non-English major classes, EFL teachers feel less motivated to develop language materials for these classes. UST-S-E 1's illustrates this relationship in her response. By the same token, UST-S-WE 1 emphasizes that her students do not want to attend instructional activities, and this reduces her willingness to bring language materials to the 12th grades. Considering EFL teachers' feelings, it can be inferred from UST-S-E 2's response that EFL teachers are affected by the students' approach towards the lesson.

"There is no motivation [to learn English] in the 12th grades [in the science upper secondary schools]. So, I do not want to prepare any materials to bring into the classes." [UST-S-E-1]

"The 12th graders want to take the tests of other courses [in English language lessons. When you teach something, they do not give attention [to the content]. After a while, you give up." [UST-A-WE-1]

"As English language lesson is considered worthless [in the 12th grades], we feel worthless, as well... This unavoidably changes our approach towards the lesson." [UST-S-E-2]

In non-English major classes, BCT and FCT preparations inhibit language learning practices. The fact that English language is not covered in high-stakes test specifications causes loss of motivation for both teachers and students.

Washback Effects of FLT on Language Teaching and Learning

Methodological Shift in Language Teaching

In the 9th and 10th grades, EFL teachers design their instructional practices within the framework of teaching four language skills. More specifically, UST-A-E 1 provides an insight about how she teaches four skills in an integrated way. She uses songs for listening activities and expects her students to produce topic-related written outputs. In a similar vein, UST-A-WE 4 and UST-S-E 2 specify the language skill which to they give more priorities in the 9th and 10th grade. According to their responses, speaking and listening skills are

widely practiced in these grades. In consonance with the ELT curriculum, UST-A-E 2 utilizes technology at the beginning levels of upper secondary education. For example, she brings short videos with subtitles into the classroom, and they practice listening skills there. She also considers the students' interests and needs in the selection of language materials.

"It could be a general listening or dialogue. Sometimes I use a song as an additional [task]. For example, [I employ the activities of] filling in the blanks [in the lyrics] or [finding] the main idea given in the song. Then I make a transition to speaking [skills] ... At end of the unit, I expect them to write. For instance, if the topic is about comparison, I ask them to make comparisons."

[UST-A-E-1]

"In the 9th and 10th grades, we try to enhance [the students'] vocabulary, enable students to understand what they listen to and give simple spoken responses." **[UST-A-WE-4]**

"I can focus on speaking in the 9th and 10th grades. Speaking is the skill to which I give most importance. I do speaking and listening practices." **[UST-S-E-2]**

"I try to find the topics that may get their attention like writing subtitles for short videos. In this way, we can integrate technology." **[UST-A-E-2]**

Whereas EFL teachers are in attempt to teach four skills in the 9th and 10th grades, these practices radically change in the 11th and 12th grades regardless of the class type like language or non-English major classes. UST-A-WE 4 summarizes this methodological shift gradually in his response given below. UST-A-WE 4 mentions two different perspectives, from which English language teaching practices differ. In the first one (non-English major classes), teaching focus is on speaking English language in the 11th grades. On the contrary, in language classes, teaching focus changes, and grammar gains prominence in the 11th grades. That is to say, the students in language classes begin practicing grammar topics in the 11th grades.

"At the beginning of the 11th grades, there are two ways to follow... In non-major English classes, we try to have students speak in target language. As for language classes, we start from the scratch in English grammar: from pronouns, tenses to modals. Whereas we have time for speaking in other non-major English classes [4 lesson hours per week], we only teach to test in language classes [16 lesson hours per week] While English language lesson depends on language objectives in the 9th and 10th grades, there is a transition from language objectives to [grammar] topics in the 11th and 12th [language] classes. I hope that I could show the discrepancy here." **[UST-A-WE-4]**

UST-A-WE 3 agrees on UST-A-WE 4's opinion about the fact that an instructional change occurs in the 11th grade. He puts emphasis on a contradictory in language classes between what is taught and what should be taught. In fact, he asserts that this change is totally against his teaching beliefs. Nevertheless, he feels the necessity of teaching to tests because of high-stakes examination.

"There is a sharp contrast here. This is not what I plan to teach. I want the students to [communicate in the target language] because I think that language is the life itself, not a lesson. [However] in language classes, we teach what is to be tested in FLT. To be honest, I am not happy with it but university entrance exam drives us to teach in this way. When I teach listening and speaking here, students will make objections [on these practices] because what I have taught is not measured in FLT. They are right. There are so many grammar topics to learn but not much time. They also have to take so many language tests." [UST-A-WE-3]

In summary, skill-based language teaching can be practiced in the 9th and 10th grades through technology integration. However, in language classes, exam-oriented teaching takes place because all the four skills are not included in FLT's coverage. Obviously, FLT shapes language teaching practices and causes a methodological shift in upper secondary education.

Teaching Test Strategies

The previous section discusses that language teaching becomes more mechanical and exam-oriented in language classes because of FLT washback. During this exam-driven teaching, UST-A-WE 3 observes a mismatch between the nature of foreign language teaching and his teaching practices. As an example of his practices in language classes, they analyze the previous language tests by giving focal attention to the coverage of test content. In a similar way, UST-A-WE 6 addresses how she introduces a sample language test to 11th graders. She conducts an in-depth analysis on the item types and frames what skills and strategies FLT takers should get mastery in.

"Language classes are the classes in which I have the most difficulty [in teaching] and do not think that I teach [English language there]. EFL teachers should focus on [foreign] language in these classes but ironically we focus on the [centralized exam] because the students will take this examination. We examine the past exams in terms of [content] coverage." [UST-A-WE-3]

“At the beginning of the term, I brought a short test into the classroom and introduced the item types like sentence completion and cloze test. I explained the importance of enhancing vocabulary.” [UST-A-WE-6]

“In the language classes, we have strategy-based training and this prevents us from looking at the whole picture” [UST-A-WE-2]

UST-A-WE 2 also reveals that he offers training on test strategies but this training hinders focusing on language learning as a whole. Put together, EFL teacher responses indicate that high-stakes centralized test drives them to get their students familiarized with item types and test strategies.

Mastery of Language Skills

The findings related to low-stakes assessment shows that washback facilitates spoken production. By comparison with low-stakes assessment, FLT debilitates the attainment of speaking skills as underscored by UST-A-WE 2. He believes that this high-stakes test improves reading comprehension and vocabulary because it includes well-structured items. UST-A-WE 2 brings foreign language teaching into question in language classes and notes that they practice basically reading comprehension that is largely measured in FLT. UST-A-E 2's response draws attention to similar findings. Her students practice reading comprehension more because of the FLR coverage. Hence, this skill is enhanced in language classes.

“Here [in language classes], do we really teach a [foreign] language? Yes, they learn something for the sake of the exam, at least reading comprehension. We definitely focus on reading as the centralized exam is reading-based.” [UST-A-WE-3]

“At macro level, it hinders [the development of] speaking skills since they are not tested in FLT. Likewise, it is the same for writing skills. However, I think that it supports learning vocabulary at micro level. I consider the items of reading comprehension well-developed. It measures the knowledge of text-structure, discourse organization, inference and implications. In general, they are well-developed to me.” [UST-A-WE-2]

“They put extra effort into [learning] reading comprehension since [this skill] is tested [in FLT]. They need to improve vocabulary and understand the main ideas of the reading passages. They feel this necessity because these [points] are tested.” [UST-A-E-2]

FLT washback on mastery of language skills is observed in the following university-level education. Field Expert 8 emphasizes the long-lasting influence of FLT even after administration. Test takers, who pass the exam, may choose to be educated at faculty of education and become a language teacher. Field Expert 8 makes an investigation on their mastery of language skills and surprisingly explains that those being successful in FLT are not able to perform four language skills effectively because of the centralized exam specifications. She also foregrounds that the change in testing system is to sort these problems out.

“[The student teachers] come with very high scores on the FLT exam, I thought they would say, they would be able to speak, they would be able to write... you will see that the results are really disappointing, they would be telling me that they are successful or they think they know English because they were able to answer, let’s say 73, or 74 out of the 85 questions on the language section of the exam.” [Field Expert 8]

In conclusion, FLT has an obvious influence on improvement of reading skills and vocabulary enhancement. Other skills are not largely practiced in language classes since FLT test specifications are not based those skills, which is extensively discussed in the section entitled “*High-Stakes Test Blueprints*”. It is also inferred from the expert response that this influence could be long-lasting and noticed in upper levels of education.

Washback Effects of FLT on Policy-Making

Teacher Designation

Another washback effect of FLT is based on teacher-related decisions. UST-A-WE 3 demonstrates how language teachers are designated to the teaching grades in upper secondary education. In the school communities, most stakeholders concentrate on exam-preparations, and as a policy of these preparations, language teachers who are experienced in preparing the students for the high-stakes exam are mostly appointed to the 12th grades in order to share their expertise with the students.

“[English language] teachers are designated to some [certain] grades in line with the centralized exams results. For example, there are some teachers in the district who are famous for and experienced in exam-preparation. These experienced teachers are assigned to the 12th grades to teach [English].” [UST-A-WE-3]

“Before being appointed to MoNE, I used to work at private exam-preparation courses. As I worked there for 6-7 years and have experience in exam preparation, the school administration has assigned me to the 11th grades to teach [English]. Since the beginning of the term, I have done practices and preparations for only high-stakes test.” [UST-A-WE-6]

UST-A-WE 6 gives a personal example in relation to teacher designation. Before MoNE, she used to work at a private exam-preparation course. Owing to this teaching background, school administrators assigned her to (FLT) language classes. She only does exam-related practices there. Consequently, upper secondary schools develop some policies for FLT, one of which is teacher designation. It means assigning language teachers experienced in high-stakes exam to FLT exam-preparation classes in order to do exam-oriented practices.

Washback Effects of FLT on Language Assessment

Methodological Shift in Language Assessment

EFL teacher responses indicate that low-stakes language assessment has undertaken a gradual shift from skills-based and technology-integrated methods to more exam-oriented ones through the exam-preparation grades. At the beginning levels of upper secondary education, language teachers administer alternative and technology-based assessment in addition to formal examination. Especially in the 9th grades, UST-A-WE 3 integrates technology into assessment and employs e-portfolios to keep track of student work. He also observes his students’ in-class performance. Similarly, UST-A-WE 5 utilizes rubrics for portfolio and project-based assessment and UST-A-WE 2 employs portfolio assessment for writing skills.

“It changes according to dynamics of the 9th graders but I try to put emphasis on formative assessment [process-oriented assessment]. I use e-portfolio for this formative assessment. I ask the students to keep their products and I use the technology there. Besides, I administer the quizzes as a part of formative assessment. Our performance-based grades are based on [in-class] observation. I have some criteria in grading the students’ performance.” [UST-A-WE-3]

“We have used portfolio assessment, projects, in-class performance assessment along with rubrics.” [UST-A-WE-5]

“We can regard their writing assignments a portfolio assessment because they keep them and then edit after I give whole-class feedback. They submit me the second papers that are edited.”

[UST-A-WE-2]

As for the formal written exams, UST-A-VE 1 highlights below that mechanical aspects of language including grammar are measured in a contextualized way in these exams. Plus, UST-A-E 1 develops a variety of item types to measure the basic reading skills and vocabulary in these exams. Teachers responses elucidate that performance-based and written exams were conducted in the 9th and 10th grades in addition to alternative assessment through portfolios, e-portfolios, projects and assessment tasks.

“We assess mechanical knowledge in the written exams. They include the grammar topics we have learnt throughout the year. We ask them in context. Rather than language skills, we assess grammar in these written exams.” **[UST-V-E-1]**

“We give a place to a reading passage in the written exams and ask questions related to this reading passage and also assess vocabulary within it. [The question types are] question and answer, true and false, open-ended questions, and matching.” **[UST-A-E-1]**

After the students have completed the 10th grade, they decide on the field (i.e. non-English major classes like sciences or social sciences and English language classes), in which they will cover and practice more field-based subjects. UST-S-E 1 makes a point about assessment procedures in the non-English major classes and states that the difficulty level is relatively low in these exams. UST-A-E 3 makes a confirmation on the difficulty level of the assessment although they attempt to administer skills-based performance assessment.

“In the 9th and 10th grades, we conduct the exams including [all the skills] listening, speaking, writing and reading. In the 12th grade, we prepare the exams including four skills, too but they have low difficulty [since] everybody focuses on [the centralized] exam and do not give much attention to the school subjects.” **[UST-S-E-1]**

“We assess [their language performance] but the difficulty level of the exams is too low. We certainly carry out performance-based assessment [skill-based assessment]. [For example], reading passage is as short as the students can do. I read aloud the listening text. We give them some topics for writing in advance. We assess them but do not push the students [with difficult questions]. **[UST-A-E-3]**

“Normally, if you have B2 level students, you get them to reach C1 level. However, we use A1 or A2 level questions. Since the students get prepared for [the university exam], we do not affect their diploma scores.” [UST-A-WE-4]

UST-A-WE 4 elaborates on this difficulty issue and explains the reasons behind the low difficulty. According to his response, the 12th graders are expected to get mastery in English language at C1 level; nonetheless, EFL teachers use A1 or A2 level exam items so that the students could get higher diploma scores. It means that performance-based assessment is still implemented in the 11th and 12th grades; however, the difficulty level of this assessment is not as expected in the ELT curriculum.

Exam-Oriented Language Assessment

Whereas skills-based performance assessment is conducted in the 9th, 10th and non-English major 11th and 12th grades, EFL teachers do not administer such an assessment in language classes. Instead, a sample of FLT is prepared and presented to language learners as indicated in UST-A-WE 4's response below. He says that they take into consideration FLT item types in the development of teacher-based written exams. A similar exam-preparation procedure is closely followed by UST-A-WE 6 who fundamentally assess test skills targeting at reading comprehension and sentence completion.

“I think it is better to use a language exam equivalent to [the centralized] exam in language classes. We prepare the exams that are appropriate to the level and item types of [the centralized] exam.” [UST-A-WE-4]

“I would prepare an exam that is similar to [the centralized] exam with 80 items and includes the vocabulary, grammar, sentence completion and reading comprehension.” [UST-A-WE-6]

“I used to prepare weekly vocabulary quizzes while teaching in the 11th grade.... Vocabulary is very important for those who take language examination. We prepare exams but these exams do not include listening and speaking skills because we have to prepare these students for [university entrance] exam.... In the 12th grade, we used multiple-choice items in the written exams.” [UST-A-WE-1]

UST-A-WE 1 employs vocabulary quizzes and multiple choice-based written exams in language classes but does not assess listening and speaking skills due to the centralized test specifications. Besides the quizzes and written exams, UST-A-E 2 administers mock tests and then they discuss the answers one by one.

“Especially in the 12th grades, I administer mock exams. At the end of the examination, I hand out the answer keys, and they check the answers. We discuss the items one by one.” [UST-A-E-2]

In simple terms, whereas BCT and FLT bring about a shift in the 11th and 12th graders’ assessment practices, more exam-driven language assessment is adopted in language classes.

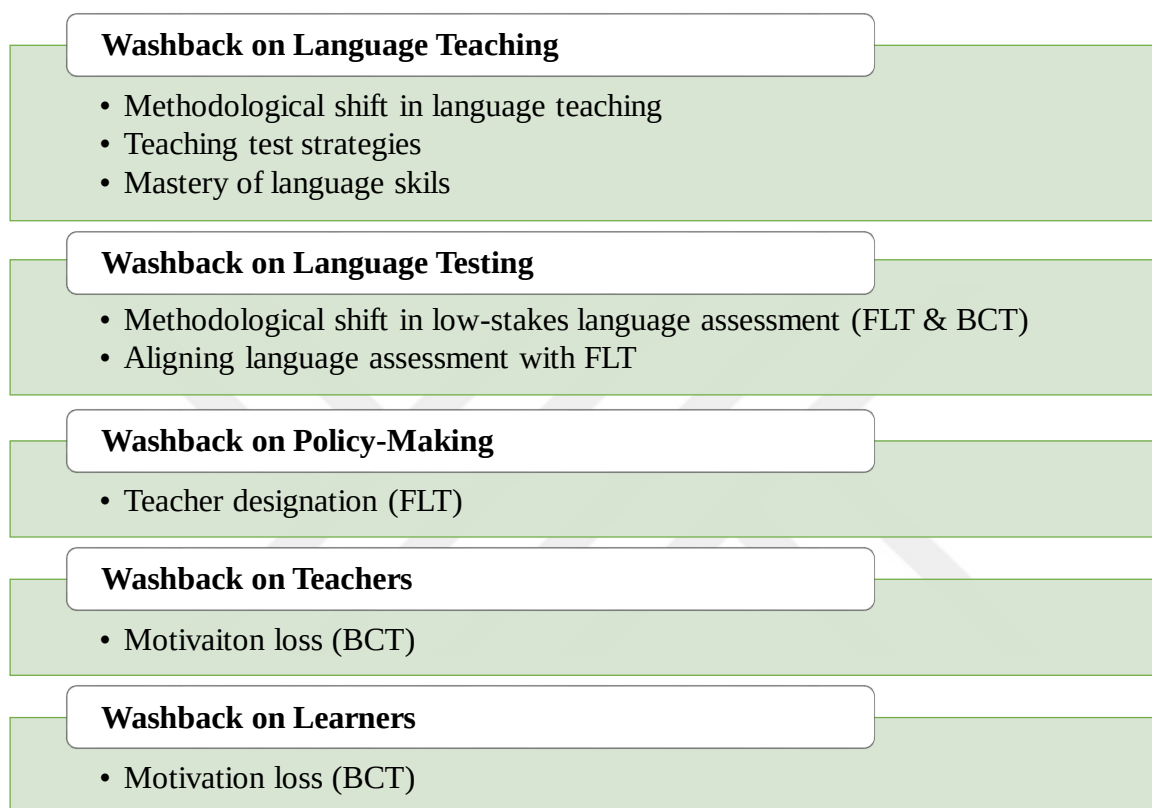


Figure 19. Washback effects of FLT and BCT and language teaching, learning, testing, learners and teachers

Figure 19 holistically depicts how high-stakes tests influence English language teaching dynamics. BCT affects language instruction in terms of teacher and learner motivation. As for FLT, strategy training, mastery of some certain skills and teacher-related policies are driven by this high-stakes test. A methodological shift is also observed in both teaching and assessment. Figure 20 presents a data-driven framework of this methodological shift from the 9th grade to 12th grade. Upper part of the framework reflects upon teaching practices while lower side visualizes assessment practices.

UPPER SECONDARY EDUCATION

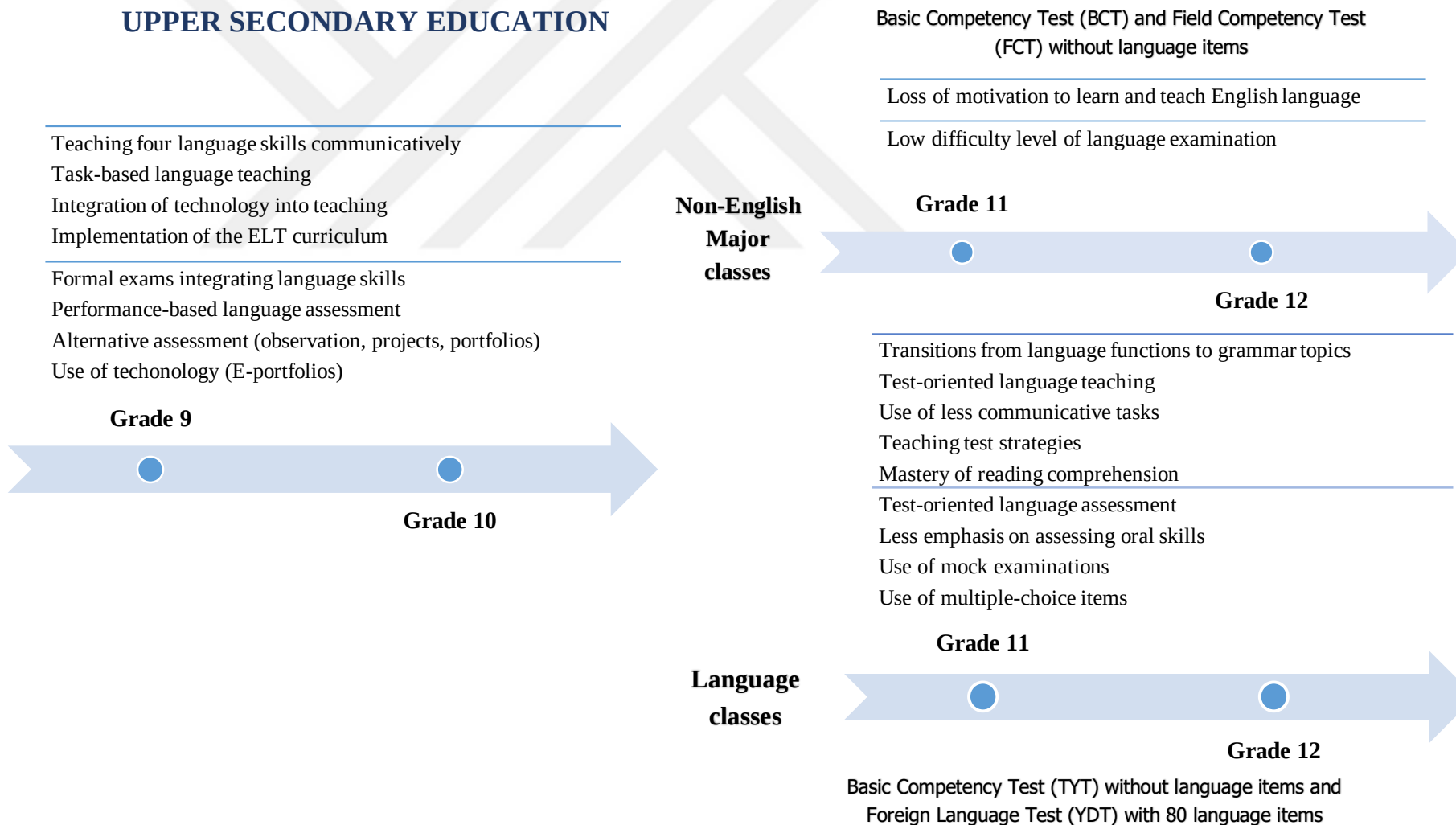


Figure 20. English language teaching and testing practices in upper secondary education: Radical shift from the grade 9-10 to the grades 11-12

Triggering Factors for Washback Effects of High-Stakes Tests

Stakeholders' Perceptions and Expectations

School administrators, students and parents have some expectations from EFL teachers with regard to language teaching and testing. For instance, UST-A-WE 5 points out that language teachers are expected to give the students opportunities to do practices in line with the centralized exam. Likewise, the parents believe that their children should not be forced to learn English more particularly in the 12th grades as highlighted in UST-S-E 2.

“School administration expects us not to push the 12th graders too much [to practice English] because [they] focus on the [centralized] exam.” [UST-A-WE-5]

“Our parents, especially in the 12th grades, want us not to push their children [to learn English language].” [UST-S-E-2]

Similar to the administrators and parents' expectations, students concentrate on exam-oriented testing. To take an example, UST-A-WE 2 mentions that the students ask about why they are not tested as the centralized exam does. It shows that their expectations are mostly related to use of tests in assessment. In this respect, parents believe that their students should take many language tests in order to keep up with others as given in UST-A-WE 2's response.

“At the end of performance-based assessment [language classes], when the students had low scores in listening, they wanted me to cancel the exams and asked me why I do not use language tests in the exams.” [UST-A-WE-2]

“One of the parents whose child changed department in the 12th grade came to the school and asked me whether he needed to take many language tests to catch up with his friends.” [UST-A-WE-2]

Field Expert 1 justifies the rationale behind why the stakeholders value these exam-oriented teaching and testing. He argues that their learning background shapes their perspectives. From a different point of view, Field Expert 4 puts forth the university examination as a main motive that attracts the stakeholders' attention to practice or mock tests. According to the stakeholders, practicing the exam means taking tests as highlighted in Field Expert 4's response. Consequently, both students and parents expect to focus on test taking in language classes rather than communicative activities.

“This is normal because this is the way they were taught and they were tested when they were students.” [Field Expert 1]

“They think that their children should study at university. In order to study at university, they need to pass university entrance exam. In order to pass the university exam, they need to practice language tests and get prepared for the exam at school. Therefore, all educational theories collapse there. It becomes useless to employ games, have fun, teach in a stress-free atmosphere. It turns into how many correct answers they have in mock exams.” [Field Expert 4]

In this respect, Field Expert 8 pays attention to the sense of achievement in the society. The stakeholders attribute learners’ achievement to the centralized exam results. They develop an understanding that those passing the centralized exam are successful in language learning. This perspective is also validated in Field Expert 5’s response given below. Accordingly, the sense of achievement directs the stakeholders’ motivation to practice the centralized exam itself in teaching and testing. Field Expert 8 states that the stakeholders will focus on alternative ways of assessment when they are measured in this way in the centralized exam.

“The exams, of course, what is considered the success? ... Because of how we treat exams, how we define success? If you do not have exams, multiple choice exams, we do not prepare for them, what do you do, you put time and effort to work on your projects.” [Field Expert 8]

“In our educational culture, the exam is always very important because exams and grades are always regarded as the evidence for success.” [Field Expert 5]

Briefly, the stakeholders expect EFL teachers not to be severe in language teaching in the non-major English classes. Also, parents and students give importance to the use of language tests in language teaching and learning because the centralized exam is test-focused and this specification shapes the perceptions of achievement.

Socio-Cultural Dynamics and Employability Policies

The centralized exam has some social, economic and cultural consequences, one of which is illustrated by UST-A-WE 2. He talks about his dialogue with the students about the university education. According to the response, the students perceive the tertiary-level education as the way of going out of the city they live in. At this point, the university entrance exam mediates the social and cultural changes in the students’ life.

“I asked my students at which university and why they wanted to study. Most of the students answered that they were planning to change the city they lived in. They view the exam as a way of going out of the town.” [UST-A-WE-2]

In the previous section, it is elucidated that the stakeholders view the university entrance exam as the indicator of the achievement. UST-A-WE 3 elaborates on how these perceptions are shaped, and he claims that the centralized exam largely influences the parents and students’ life. In other words, this exam is considered a gateway for all important decisions. UST-A-WE 5 agrees on this response and explains that the centralized exam determines important decisions in life.

“In our country, passing the university entrance exam is regarded as an indicator of success... Parents feel proud of their children’s success. They say that their own children have achieved something and have got these scores. What is the reason for [these perceptions]? Because our criterion for success depends on university entrance exam... We suppose that we get everything in life by passing university entrance exam.” [UST-A-WE-3]

“University entrance exam is the life, everything, the determiner of the life.” [UST-A-WE-5]

To be more specific, one of the decisions that are determined by the centralized exam is the profession as emphasized in Field Expert 6’s response. This result is also confirmed by UST-A-E 1 and UST-A-E 3 who assert that university education means getting a job and making money, which is considered the financial outcome of university-level education. UST-A-WE 1 comments on this outcome within the contextual setting of university education. When the university education is seen as a way to get a job and earn money in a country, university entrance exams will gain prominence in this society.

“There is a university entrance exam. Everyone gets prepared for this exam with the purpose of getting a job and better life. That is why the exam itself is effective [in shaping the stakeholders’ perspectives].” [Field Expert 6]

“In our country, university is the gatekeeper for getting a good job and earning good money.” [UST-A-E-1]

“Students study at universities with the aim of getting a job. Their purpose is to pass the university exam, to have a good job and a better life.” [UST-A-E-3]

“Unfortunately, the society’s expectation from education is to make money in our country. The parents expect their children to graduate from the university and get a job. That is, they regard the university [education] as the step [that leads to] making a living.” [UST-A-WE-1]

Field Expert 9 opens a discussion about this financial aspect of university education and entrance exams. He wonders about whether or not university entrance exams will be so significant when the stakeholders have no concern about the employment or affording their life.

“What happens if the individuals are employed or they may afford their life even though they are not university-level graduates.” [Field Expert 9]

This question is fully addressed in UST-A-WE 1 and UST-V-E 1’s responses. Another dynamic that mediates the washback effects of BCT and FLT is the employability. For instance, the exam-oriented teaching and testing are not observed in some vocational schools because they need to learn English language because it is the requirement for being employed. UST-A-WE 1 gives a specific example and accentuates that her students were motivated to learn English because of their professional experiences. So, the centralized exam influence is not traced in these types of school.

“It [English language learning] was important when I worked in Anatolian Technical Upper Secondary School. Students were going to become plane technician. They started working in Turkish airlines or other private companies. English language was important for these students. There was no employability for them without English language.” [UST-A-WE-1]

Similarly, UST-V-E 1 indicates that there is no effect of the centralized exams on both teaching and testing because the students will be employed as soon as they graduate from the vocational upper secondary schools. In addition, they practice English language during their internship; for this reason, they take part in language practices in the classroom. Not only language teaching activities but also assessment tasks are communicative in those schools as highlighted in the following response.

“This is Tourism Upper Secondary School, even if English language is not included in [the centralized exam: FCT], the 10th, 11th and 12th graders do internship and they have to speak English in order to communicate with tourists. Thus, their interest in English language never fades away; otherwise, they will not be employed, they are aware of this fact... Considering my own school, I cannot say that they have the exam anxiety because they get the job as soon as they graduate from the school... We have specific topics or activities for each grade. We give project assignments from these topics. ... For example, we had vocational English, and the students did role-playing for the topics like “Hotel Check-in” or “Serving”. These projects made considerable contributions to the grading.” [UST-V-E-1]

Field Expert 6 clarifies this issue with causal relationship between supply and demand. When the students have demands on language learning, there will be communicative activities. When they need to pass the exam, there will be exam-oriented teaching.

“There are two concepts: demand and supply. Supply is shaped by demand. In fact, this reflects on our education, economy, sociology, economy.” [Field Expert 6]

To summarize, university education takes an active role in shaping the individuals’ social, economic and cultural life. Therefore, passing the university entrance exam brings about social, economic and cultural outcomes. However, employability is a variable that mediates the effects of the centralized exams. When the students guarantee to be employed at the end of upper secondary education, some effects like exam anxiety or exam-oriented practices disappear in the school community.

Institutional Dynamics

Passing university entrance exam holds important meanings for the stakeholders. Considering the results of university entrance exams, school administrations makes various advertisements as stated in UST-A-E 3 and UST-A-WE 2’s responses. The main purpose for disseminating these advertisements is to persuade lower secondary graduates for recruitment.

“The school administration will make an advertisement that the school has enabled the students to pass the university exam and study at the faculty of medicine, law etc. It is a kind of advertisement to increase the student recruitment. In this way, [they] persuade others to apply for our school.” [UST-A-E-3]

“The school administration considers making advertisements. A few years ago, one of the students in language classes passed the exam and got into a prestigious university. This is the typical example.” [UST-A-WE-2]

UST-A-WE 5 underlines that student recruitment increases the school prestige. What’s more, the school prestige affects the following recruitments of the lower secondary education graduates as shown in **UST-V-E-1**.

“The school administration aims to enhance the statistical [output] of the school. The exam results are regarded as the basis for more successful students’ recruitment. Thus, they ensure the school prestige.” [UST-A-WE-5]

“It is a matter of prestige. It guarantees the student recruitment for the following years.” [UST-V-E-1]

In simple terms, the centralized exam results are used by the school administration to capture the students’ attention for recruitment. The more students make applications for recruitments, the more prestigious schools become. In addition, there is two-way relationship between student recruitment and school prestige. More recruitments also promote the school prestige.

High-Stakes Test Blueprints

How test specifications are defined in high-stakes assessment is a significant factor that governs washback effects on instructional settings. In relation to these specifications, Field Expert 4 argues why BCT does not measure all the subjects and content given in teaching curricula. For instance, English language skills are taught in upper secondary education but not measured in large-scale centralized test like BCT. He points out this dilemma between teaching and testing as follows.

“Students get prepared in accordance with the requirements of the centralized exam. There is only one motivation type in Turkey. Is it measured in the exam or not? What is the meaning of the inclusion of the content in the curriculum if it is not measured in the exam? It does not make any sense. English language is placed into teaching programs but not measured in high-stakes exams. If it is not that important, why is it given a place in teaching programs? If you put it into the program, you need to measure it in the exams. There is an inconsistency here...” [Field Expert 4]

Likewise, Field Expert 9 underlines the fact that there will be no motivation to learn English if it is not measured. Hence, test specifications guide test takers about what to study and what to make preparations for.

“You cannot motivate [the students] to learn the content that is not measured in the exam.” [Field Expert 9]

When FLT is examined in terms of test specifications, it is seen that there is a discrepancy between ELT curriculum and large-scale test as given in UST-A-WE 2’s response. He emphasizes that the ELT curriculum does not prepare the students for the centralized exam. UST-A-WE 3 states that this centralized language exam is not related to language teaching

policies at university as well as curricular principles in upper secondary education. Since there is a mismatch between curricular objectives and FCT blueprints, EFL teachers are driven to conduct test-oriented language teaching as stressed in Field Expert 1's response.

"Do you think that ELT curriculum gets the students prepared for the [centralized examination]? No it does not. I have to adapt it. One of my colleagues says that he cannot find the topics that are covered in the [centralized] exam." [UST-A-WE-2]

"Interestingly, high-stakes exam, FLT, has no connection with neither the ELT curriculum in MoNE nor upcoming university-level education." [UST-A-WE-3]

"You see, in fact, there is a discrepancy between [testing] policies and MoNE policies. So there is a clash, mismatch. As a result of this, teachers prepare students for these high-stakes exam."

[Field Expert 1]

This issue is also discussed by UST-A-WE 5 who states that test takers in other subject-fields are aware of what content is tested; however, language learners are not familiar with which objectives will be measured.

"There are teaching programs in this country. The students are familiar with the [other subjects'] objectives that are measured in FCT. Content is clear there. What geometry formula will be tested is [explicitly] known. Except for [these topics], nothing [that could be additional or new] is measured there. It means that the exam and curriculum completely match. This is not the case in FLT." [UST-A-WE-5]

Field Expert 7 asserts that English language teaching will be performed successfully when there is a matching between the curricular objectives and test specifications. Similarly, Field Experts 2 underscores the need for change in testing system to keep the balance between teaching syllabus and testing methodology.

"It will be successful when there is a match between what you teach in the school and what is measured in those exams." **[Field Expert 7]**

"There is no correlation between outcomes of the syllabus and the testing system. The first thing we should do is to establish this harmony between the outcomes of our syllabus, our curriculum and the testing system itself." **[Field Expert 2]**

Field Expert 3 proposes that language learning will be represented as a whole by the time this connection is ensured between language teaching and testing policies. The change in testing system and inclusion of four skills will motivate EFL teachers to teach four language

skills as understood from Field Expert 1's response. These perspectives are also validated by Field Expert 8. She expresses that nothing will be improved in English language teaching unless English language is tested in compliance with teaching curriculum.

"As long as we measure language skills, we can come to the conclusion that we are having the whole picture of language learning... Obviously the picture would be incomplete without exploring, without assessing listening, writing, speaking." **[Field Expert 3]**

"You act accordingly in line with the exams presented to you. You are assessed and evaluated in line with this exam. If you do not administer [a valid] exam, you cannot get [the valid] outcomes. The first thing to do is to change exam type, they should include, cover four skills. If you automatically change, I am sure teacher will comply with the exam items. That is the power of exams." **[Field Expert 1]**

"Testing always lies behind but now we have to realize that if we do not the test as we teach, then all the effort the people put to test in the teaching, the time, the energy, the materials, the triangle of those things. Excuse me but it will go down and drain if we do not test as it is required." **[Field Expert 8]**

Suffice it to say that integration of skills-based assessment into high-stakes examination will change the intensity and value of washback effects. Considering the reasons for which skills-based assessment is not included in the current large-scale centralized assessment, Field Expert 4 gets attention to some significant dimensions. For example, MSPC must be transparent in its testing policies and is expected to publicize the centralized test items after the administration.

"In Turkey, such public exam [results] are open to filing a suit. People can file a suit against the items, let's say that they think that the 15th item is not properly prepared and file a suit against this item. MSPC has to share the questions/items in public that's why they could not prepare an item pool. MSPC has to be accountable for [the item development, test results]." **[Field Expert 4]**

In addition to transparency dimension, MSPC should be accountable for uses of test results as discussed in Field Expert 4's response below. BCT, FCT and FLT are the tests that have important consequences for the stakeholders. In such tests with high stakes, MSCP becomes accountable when the examination fails in achieving its purposes in terms of interpretation of exam scores. In order to ensure this consequential validity, objectivity should be ensured

at first. Field Expert 6 explains that there is no objection to the objectivity of multiple-choice examination. That is why language is tested with multiple-choice items.

“Why do not we administer open-ended questions? As teacher trainers, we believe that open-ended questions are better but we are aware of the difficulty to score open-ended products. In an exam which millions of people take, who will rate the open-ended answers? There is a practicality issue. In addition, the fact that there are some doubts about the scoring of an item will cause chaos in Turkey because this exam has high-stakes consequences”. [Field Expert 4]

“There is a concern of objectivity. In multiple choice items, the correct answer is obvious.” [Field Expert 6]

Another dimension that affects skills-based language assessment is practicality as mentioned in Field Expert’s response above. In FLT setting, millions of people take this examination. Field Expert 6, Field Expert 9 and Field Expert 8 touch upon the practicality issues in administrating, scoring and reporting skills-based assessment.

“Let’s say that millions of test takers have written one paragraph. At least two raters should score it. It will take time.” [Field Expert 9]

“What I think the reason is because we have a huge population in Turkey. ... The result of those exams have to be announced pretty quickly. So, this is one of the reasons.” [Field Expert 8]

“It is because there is a large number of test takers. This makes the measurement difficult. You can conduct an objective assessment but it takes time.” [Field Expert 6]

Along with the practicality of language testing due to the high population of test-takers, subjectivity concern is another debilitating factor that hinders skills-based assessment. Field Expert 5 puts emphasis on the need for rater source and subjectivity that may occur in assessment of four skills. Similarly, Field Expert 4 sheds light upon the inter-rater reliability concern. So, it means that reliability should be ensured in skills-based assessment.

“Most probably, it is because of the high population of the test takers. It is because there could be subjectivity in assessing language skills in the centralized exams and there could not be the required facilities to assess [those skills]. It requires the rater source. Besides, it is necessary to train these raters. Not everyone can rate [those skills], we need policies in this field. There should be rubrics that are publicly presented.” [Field Expert 5]

“Just imagine that thousands of people take speaking exam. In order to assess the speaking skills, we need raters and we need to ensure inter-rater reliability. It is easy to say but difficult to do.” [Field Expert 4]

Finally, Field Expert 9 foregrounds the infrastructure facilities of language testing. FLT has high-stakes consequences for the stakeholders, and each measure should be taken in advance in order to complete the exams properly across the country.

“There should not be made any mistakes in the [centralized] exam that millions of people take. Just imagine what happens if there is a cut in electricity. The exam is cancelled. [In order to avoid such problems], you can take measures like use of batteries.” [Field Expert 9]

In a nutshell, test specifications play an essential role in exerting washback effects of the centralized tests. If they require the assessment of four language skills, washback will focus on practicing these four skills. In FLT, skills-based assessment is not included. In order for a high-stakes exam to cover four language skills, LASBLA dimensions should be ensured respectively as pointed in Figure 22.

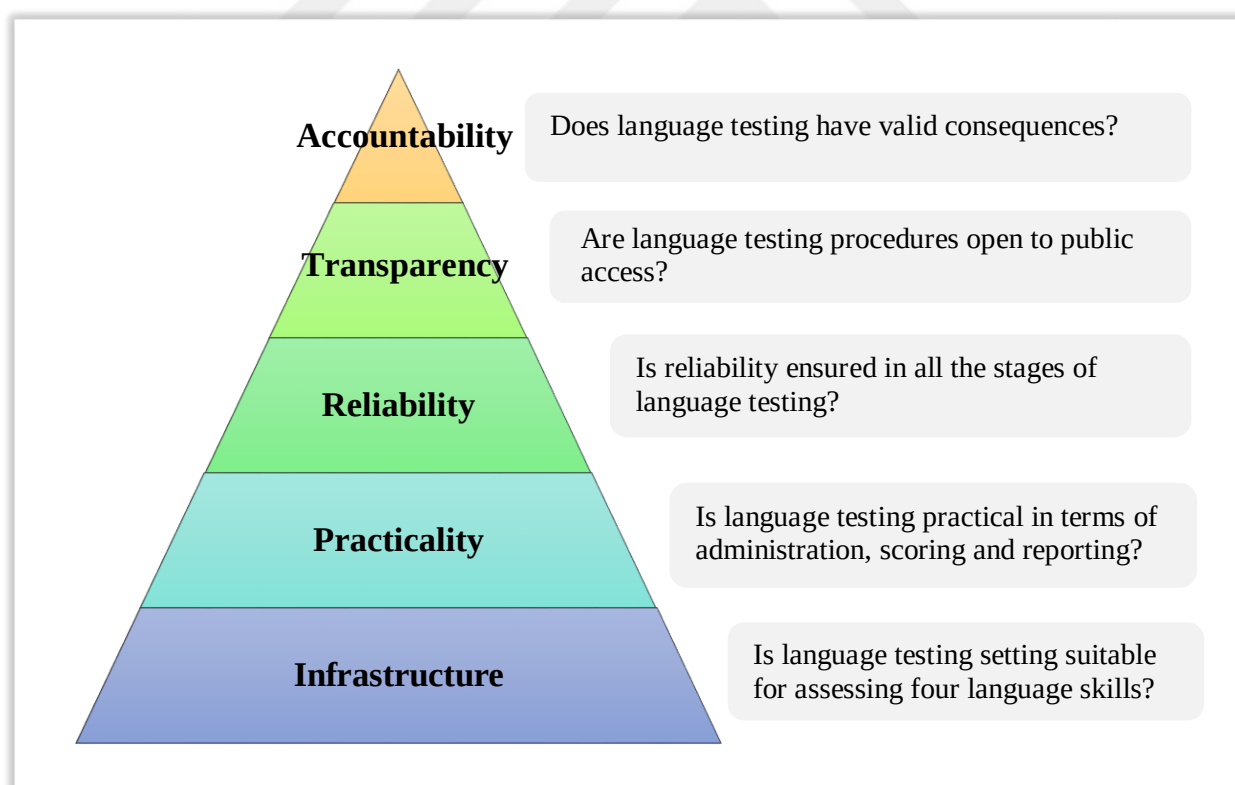


Figure 21. LASBLA: Pyramid of large-scale skills-based language assessment

Teacher Beliefs, Practices and Assessment Literacy

In the section of washback effects of high-stakes assessment, it is explicated that the centralized exams bring about a shift in EFL teachers' teaching and testing methodology and practices. Teachers have a tendency to align their practices with high-stakes test specifications through the 12th grades. However, UST-A-E 2 and UST-A-WE 5's additional non-exam oriented activities gather attention when compared to the study findings. Even though FLT specifications do not consist of assessing four language skills, these teachers employ skills-based communicative tasks in language classes.

"We use different types of questions in language classes. We ask them to write an essay rather than paragraph or give them 5 minutes to speak [in the exams] instead of [allocated] two minutes in other classes." [UST-A-E-2]

"I [observed] that the students [in the 12th grade – language classes] have weaknesses in listening and speaking as an outcome of exam preparation. Thus I brought the songs and live news, and I tried to present authentic materials. We taught speaking and listening and exposed our students [to writing] by providing samples to construct background information. As for the examination, I assessed speaking skills and showed my rubrics to them. On the other side, I conducted written exams similar to the exam format [FLT] including cloze test, paragraph, irrelevant sentence at the level of grammatical, semantic or lexical level." [UST-A-WE-5]

In UST-A-WE 5's response above, he observes that the centralized exam impedes the learners' speaking development. So, he brings communication-triggering supplementary materials into the classroom. Moreover, he prioritizes the assessment of productive skills and develops written exams including various item types. On the other hand, UST-A-WE 2 attracts attention to teachers' knowledge of testing and interprets his observation that use of test has become a tradition among language teachers. His response also demonstrates his effort to increase teacher awareness on alternative or other kinds of assessment.

"[In our teacher trainings] I observed that teachers generally focus on the technological tool to develop test items that they will use in assessment and evaluation. There is a 'test' tradition like university entrance test. They want to use multiple choice items. However, I [explain] that test is not only way of assessment and evaluation. I tell them to use [Web 2.0] tool as a portfolio [assessment] or assess the students' collaboration or interaction when they communicate with each other, or observe them and then give feedback." [UST-A-WE-3]

As stated in the section of methodological shifts, teachers tend to use exam-oriented activities; by comparison, some teachers support these activities with more communicative tasks. That proves that assessment practices may vary according to teacher preferences. Some whereas the high-stakes test specifications. It can be deduced from UST-A-WE 3's response that assessment literacy plays a significant role in mediating the washback effects of the centralized exams. Along with teachers' assessment literacy, teacher beliefs shape washback mechanisms in language teaching. For example, UST-A-WE 2 bases English language teaching on his truths as well as learning gains he obtained in pre-service education. Similarly, UST-A-E 2 states that she will continue teaching in light of her beliefs even if she is viewed as an unsuccessful teacher.

"I take into account my [pre-service education] rather than these expectations. I consider whether I have measured language objectives in the curriculum or not... I follow my own principles [because] I do what I believe [according to my own truths]" [UST-A-WE-2]

"It does not matter that I am regarded as a successful teacher or not. If everything is clear in my conscience, there is no problem for me even though I am judged as an unsuccessful teacher."
[UST-A-E-2]

So far, it has been discussed that teacher practices may show differences even though the centralized exam dominates the instruction, and assessment literacy and teaching beliefs are two basic mediating factors that formulate washback on teaching and testing. Field Expert 3 and Field Expert 5 elucidate what shapes these teacher-related factors. According to Field Expert 3's opinions, in-service experiences highly influence their assessment beliefs and practices.

"What I would say, these assessment beliefs are very much shaped on the basis of their real experiences. Think of a student teacher, a graduate from the university full of fresh ideas, when she gets into the system, she gets into the profession, she comes across a lot of difficulties, there is a couple of ways that either she is going to stick to principles that was put forward by professors in their own education [or] she is going to go with the flow. Mostly, she has to go with the flow because assessment is not very well structured. In this particularly gut, teaching beliefs even though have a long history in relation to their own learning experiences, these assessment beliefs are largely influenced in the in-service period." **[Field Expert 3]**

This perspective that puts emphasis on in-service teaching and testing is also confirmed by Field Expert 5's response. She exemplifies the stakeholders that may affect language

teachers in terms of testing and makes a point about how novice teachers are influenced by mismatch occurred in belief systems. Concerning this influence, Field Expert 2 acknowledges the importance of in-service education about how to develop exams, assess four language skills and conduct communicative assessment because testing beliefs are largely affected in the school community.

“EFL teachers, after they graduate with fresh ideas, may be affected by their colleagues, administrators, students and parents. In most schools, the common exams are prepared by group teachers. Newly recruited teachers can be affected if there is a discrepancy in a certain belief system between experienced and novice teachers.” **[Field Expert 5]**

“The problem is that that newly appointed teachers as well as previous teachers at state schools have a system in mind. Previous teachers, let’s say, have prepared some questions for the exam. Newly recruited or newly appointed teachers try to use the same type of the question. There is a tradition among the teachers, So I believe that they are not flexible... That is why they need in-service training about testing they have to learn how to prepare questions, how to test communicatively. They should know to integrate language skills and language components in a specific exam.” **[Field Expert 2]**

Last but not least, teacher practices, beliefs and assessment literacy are the factors that lessen or increase the intensity of washback. The results show that in-service teaching and testing experiences and other parties of education are influential in shaping these factors. Through in-service training about how to conduct skills-based assessment, teachers’ assessment literacy may be improved; and this may result in more curriculum-oriented practices.

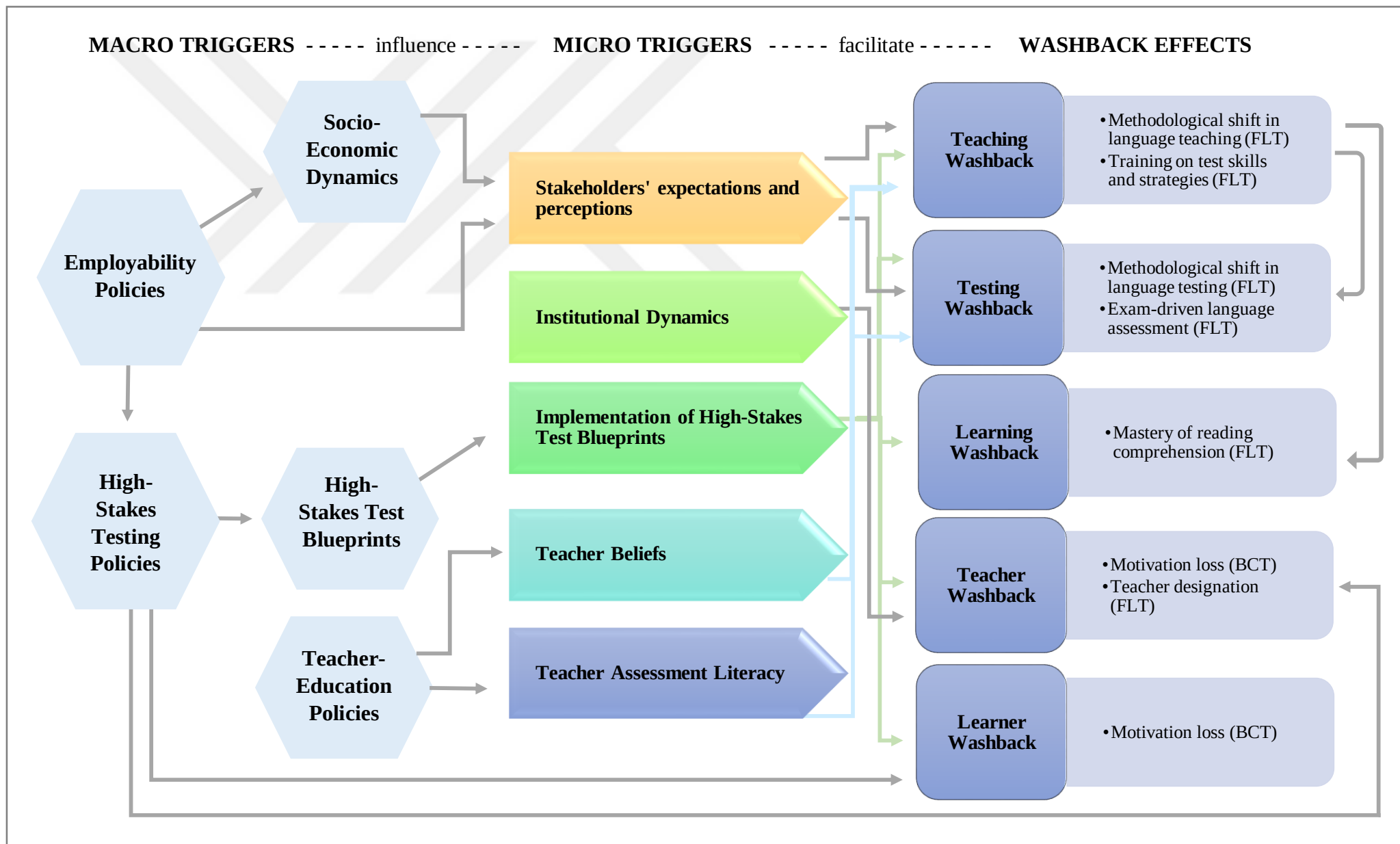


Figure 22. The triggering factors that bring about the effects of centralized large-scale examination on EFL teaching dynamics

CHAPTER V

DISCUSSION

This section includes summary and interpretations of the findings that are presented in the previous chapter. The findings will be discussed within the target case of the study with respect to the units of the analysis. In addition to the discussions and interpretations, research findings will be compared with other studies to indicate similarities and differences in relation to relevant literature.

Case 1: Primary Education

Within the primary education case, seven teachers were interviewed in total (three of them work in both primary and lower secondary education). The first result was related to the methodological shift in language assessment that is triggered by examination regulation (MoNE, 2013). While there are only monitoring practices in the 2th and 3th grade, formal written examinations take place in the 4th grade. Along with curricular density and limited teaching duration, EFL teachers give up or do not commonly use administering communicating assessment tasks and alternative assessment in this grade. However, the ELT curriculum (MoNE, 2018) puts forward skills-based assessment and alternative assessment methods besides written exams. In this respect, Sarıgöz and Fişne (2018) found out similar results by indicating a mismatch between in-class practices and curriculum requirements. As the second finding shows that some EFL teachers focus on written examination in this level of education. However, this is not the same case for all primary education teachers. There are differences among EFL teachers in terms of adopting exam-oriented teaching. This finding that displays the differences in teacher-related factors is consistent with teacher factors framed by Shih (2009), Spratt (2005) and Watanable (2004). Another finding is related to the written examination anxiety. The 4th graders are influenced affectively in this

grade because of the formal pen-and-paper exams. Şenel and Tütüniş (2011) indicated the negative effect of written exams on young learners' affective domain and explained that they caused the anxiety. The primary trigger of this affective washback effect is seen parental ambitions and expectations. Through the 4th grades, parents always want their children to have higher scores in the examinations.

From an instructional point of view, EFL teachers believe that they get information or feedback on language learning or achievement of language objectives at the end of these examinations. At this point, Madsen (1983) presents that language tests can be used as the tools to understand learning strengths and weaknesses. In addition to the formal examinations, mock exams and practice of language tests are also demanded in the primary education. There are some influential factors behind these implementations. First of all, upcoming centralized exams like USE-EE or scholarship exams motive the stakeholders to use multiple-choice items. Also, there is a perception that these students will study at lower secondary education and they need to take language tests because they are commonly used in lower secondary education. Hence, the stakeholders' expectations and perceptions on teaching and testing promotes the demand on use of tests. For example, stakeholders suppose that language teaching should be traditional and not include songs, games or communicative tasks. The important one is to pass the exams and get higher scores. While parents support this traditional teaching and testing because of the socio-economic concerns, school administrators make their schools prestigious according to the results of these large-scale tests. As highlighted by Shohamy (1993) and Shohamy et al. (1996), language tests have significant results for all the stakeholders. The current study findings point out that stakeholders including parents, principles and students have expectations from test results and thereby language teaching.

Abovementioned results are typical washback effects that are observed in primary education and triggers that exert these effects in primary education. Nevertheless, there are some teacher-related factors that mediate these washback effects. For example, teacher professional identity (i.e. teachers' role in teaching), their beliefs on how to teach communicatively and their missions and purposes to teach English may change the intensity of these effects. Ha, Tran and Tran (2021) demonstrate how teacher assessment beliefs change practical choices. Accordingly, primary school teachers' beliefs pave the way for specifying language practices and then adjusting the intensity of washback effects.

Case 2: Lower Secondary Education

The case of lower secondary education consists of two basic parts to be discussed: low-stakes assessment and high-stakes testing. As mentioned earlier, low-stakes assessment is generally viewed as classroom-based assessment, and at the end of this assessment, there are no significant outcomes that might affect the stakeholders' lives. On the other side, high-stakes testing outcomes are used to take some significant decisions on test takers. To begin with, teacher-based assessment and mock tests have some effects on language teaching, learning, teachers and learners. For instance, teacher-based assessment may cause exam anxiety among lower secondary students because a certain percentage of these assessment scores will influence diploma score that is considered in transition to upper secondary education. Whereas the research studies conducted by Kılıçkaya (2016), Toksöz and Kılıçkaya (2018) indicate the exam anxiety as washback of large-scale high-stakes tests, Hung (2012) stresses the anxiety as the effect of e-portfolio assessment. Even though it is consistent with Hung's (2012) low-stakes assessment's finding, this research emphasizes the role of diploma score in bringing about the washback effect of teacher-based assessment. Apart from negative influence, teacher-based assessment results are regarded as the evidence for learner success, and also teachers get feedback about to what extent learning has been achieved. When the washback effects of mock tests are examined, it is seen that they provide feedback on learning. However, these results make language teachers accountable for the success or failure. This finding is consistent with Tsagari (2011) in terms of teacher accountability. Also, it causes alignment of language teaching with mock tests like content specifications and strategy-based teaching. This finding is similar to Chen's (2002) finding "teaching to tests".

High-stakes testing also influences language teaching and there are some socio-economic, contextual, institutional and stakeholders-related factors that facilitate these effects. First of all, USE-EE causes methodological shift in both teaching and assessment. This finding is consistent with Göktürk-Sağlam's (2018) study but different from Wall and Alderson (1993) and Cheng (1997) elucidating that the influence of tests is only observed on teaching content but not methodology. Whereas curriculum-based activities and communicative tasks are common in the 5th and 6th grades, exam-driven teaching (i.e. focusing on the skills only testes, teaching test strategies) is employed in the 7th and 8th grades. Similarly, formative, communicative and alternative assessment are so frequently conducted in the 5th and 6th

grades, mechanical assessment practices and multiple-choice based exams are preferred in the 7th and 8th grades. It causes the use of more supplementary materials in the final grades, and this centralized testing shapes learner perspectives related to examinations. To take an example, when USE-EE takers begin studying at upper secondary education, they expect EFL teachers to focus on assessment and conduct exam-driven language testing. This indicates that USE-EE shows long-lasting influence that is observed in two different cases. Watanabe (2004) lists the dimensions of washback and explicates long-term aspects of washback. Accordingly, the research finding validates this aspect of washback dimensions.

As for institutional policy-making process, lower secondary schools develop some strategies to prepare the 8th graders for USE-EE more effectively. In particular, the successful test takers get awarded at the end of examination; EFL teachers raise the students' awareness on USE-EE and they also do exam-couching that allows teachers to follow the 8th graders' study habits. The school principals expect to use more mock exams and offer opportunities to run supplementary courses. Çakır (2017) also draws attention to similar findings in his study focusing on the same high-stakes exam. Finally, the schools develop some policies related to teachers. For example, EFL teachers are assigned to the grades of lower secondary education on the basis of their performance in USE-EE. Again, they become accountable for high-stakes test results. In regard to the triggers that exert these effects on language teaching, it is seen that there are six categories: stakeholders, socio-economic dynamics, institutional dynamics low-stakes assessment policies, teaching duration and language teachers as given in Figure 17. More specifically, stakeholders' expectations and perceptions on language teaching force EFL teachers to change their teaching and assessment methodology. The main motive behind the stakeholders' perceptions is the socio-economic factors. They expect their children to have a better life and economic wealth, and they suppose that this could be achieved only through passing the centralized exams. In a similar way, test takers think that they will get social approval when they pass the exam. From institutional perspectives, lower secondary schools aim at prestige in the district; therefore, they compare themselves with others in terms of USE-EE results. In this respect, Shohamy's (1998) critical language testing theory gains importance. USE-EE has socio-economic and socio-cultural powers; for this reason, it shapes the stakeholders' lives.

Examination regulation and allotted time for teaching are other factors that bring about the effects of more traditional teaching and assessment. Finally, there are teacher-related factors

that may trigger or mediate washback effects such as assessment literacy and professional identity. Similarly, Wang (2011) puts forward that teacher knowledge, belief and perception are influential in mediating washback effects.

Case 3: Upper Secondary Education

In upper secondary education, teacher-based assessment as low-stakes assessment and BCT and FLT as the high-stakes assessment influence language teaching to different extents. First, teacher-based assessment is conducted as the common examination as required in the regulation. Also, performance-based assessment became mandatory for language-related subjects. Then, in upper secondary education, EFL teachers develop some skills-based exams and administer them commonly in the all grades. There are some effects of this assessment on teaching, learning and teachers. Specifically, this examination gives feedback on language learning. For that reason, it becomes important for language teachers. Since speaking skills are measured there, this examination promotes the learners' spoken production. However, it may cause excitement among the learners. Sometimes, it may result in exam anxiety but it motivates the learners to learn English at the same time. From teacher aspects, it increases collaboration and interaction among language teachers; on the contrary, it hinders teacher autonomy because all the exams should be conducted commonly, and all the teachers should use the same items in their classes. There are two main influential factors that facilitate these washback effects: exam regulation and influence of these exams on diploma score that is taken into consideration in transition to university level education. In Shih's (2010) study, school policies trigger the washback effects on policy-making. The current study shows that nationwide policies also mediate the washback effects of low-stakes exams.

With respect to high-stake tests, non-English major students are affected by BCT content specifications. To be more specific, this test does not include English language items, and thus the students in these non-English major classes are not motivated to learn English. They believe that it is unnecessary to English. In this regard, scholars question language education policies. If the graduates of non-major English classes need to learn English at university, why does BCT cover these skills? If it does not cover English language, why is it placed into teaching curriculum? As there is no motivation to learn English in these classes, EFL teachers feel demotivated to teach English as well. Considering educational policies,

teaching curricula suggest four-skills assessment and also local foreign language teaching policy (MoNE, 2006) puts emphasis on four-skills teaching. However, there is not an exact match between high-stakes test specifications and language policies.

In consideration of FLT that is taken by the learners in language classes, there are some effects on teaching, assessment and teachers. For example, while the students practice four language skills in the 9th and 10th grades, they only focus on grammar topics, reading comprehension, vocabulary enhancement and test strategies as Akpınar and Çakıldere (2013) come up with similar results in a high-stakes language testing. It is also the same for language assessment in these classes. Whereas there are mainly communicative assessment tasks and alternative assessment in use in the 9th and 10th grades, assessment becomes mechanical and test-oriented in language classes. In addition to these effects on teaching and assessment, language teachers are designated to language classes with regard to their experiences in exam-preparations. Considering the triggering factors of this influence on teaching and teachers, it is concluded that there are five main types of operational forces: stakeholders, socio-cultural dynamics and employability, institutional dynamics, high-stakes test blueprints, and teacher-related factors. Firstly, stakeholders' perceptions and expectations shape what practices EFL teachers employ in both language and non-major English courses. Besides their perceptions and expectations, their sense of achievement and educational background are also influential in determining teaching and assessment procedures. The fact that the centralized exams are the gatekeepers for socio-cultural experiences, getting economic wealth and employment also increases the power of the exams. Upper secondary schools attach importance to these centralized exams because they will ensure their prestige among others, and lower secondary graduates will have more demands on recruitment. Another significant trigger here is high-stakes test blueprints. It is deduced from teacher responses that there is a mismatch between curricular underpinnings and test specifications. This causes the implementation of a hidden curriculum in language classes. Also, there are five dimensions that need to be addressed in order to assess four language skills and increase positive effects of FLT. These dimensions are infrastructure, practicality, reliability, transparency and accountability. If they are embedded in high-stakes testing policies and ensured in MSPC, curriculum-based language teaching will be promoted in language classes. In addition, upper secondary education needs more comprehensive language testing and teaching policies on which these five dimensions are reflected upon.

Similar to lower secondary education, teachers' assessment literacy, teacher beliefs, pre-service education and in-service education are among the main triggers for washback effects in upper secondary education. They may trigger or mediate the washback effects according to the power of centralized exams.





CHAPTER VI

CONCLUSION AND IMPLICATIONS

This section presents concluding remarks on the current research study and puts forward some implications for EFL teaching dynamics in order to promote the positive washback effects of language assessments and centralized tests.

Conclusion

This research aimed to present how washback operated in EFL teaching in Turkey and what macro- and micro-level triggers forced these washback effects. In order to achieve this aim, interviews were conducted with EFL teachers and field experts about the research constructs on the basis of case study research design. Qualitative data were analyzed with content analysis. The results were given under emerging categories in relation to the cases and unit of analysis. The major findings illustrate that there has been emphasis on exam-oriented teaching and assessment through the final grades of each level of education. The main triggers behind this finding are socio-economic and cultural dynamics, stakeholders' perceptions, institutional expectations and teacher-related factors. Especially in upper secondary education, language testing policies are the main influential factors. In order to promote positive washback in these cases, it is necessary to develop language testing policies that are consistent with language teaching policies. The following section depicts how washback effects operate in EFL teaching dynamics in Turkey.

How Washback Effects Operate in EFL Teaching Dynamics in Turkey

In primary education, low-stakes language assessment policies shape how classroom-based teacher assessment should be conducted. These principles enable policy-makers to formulate

examination regulations. Socio-economic factors basically related to the stakeholders' current financial status, motives for economic wealth and more prestigious life standards, influence the stakeholders' perceptions and expectations from language teaching and testing. Similarly, high-stakes language testing policies shape their perceptions and expectations. For example, test specifications including multiple-choice items attract their attention to focus on use of language test items in both teaching and learning. Obviously, these triggers drive formal examinations to bring about some effects in primary education such as methodological shift in assessment, exam-oriented teaching, exam anxiety, test-oriented assessment and demand on use of multiple-choice items. However, these effects are filtered out by teacher missions and beliefs. That is why the intensity of these washback effects are not the same in each educational setting.

As for lower secondary education, washback effects of both low-stakes assessment and high-stakes assessment are triggered by a number of factors at macro- and micro-level in Turkey. To begin with, the 8th graders take a centralized exam at the end of the academic year. In order to conduct such a high-stakes test, MoNE has high-stakes testing policies, which are influential in specifying low-stakes testing policies and shaping the stakeholders' perceptions on language teaching and testing. Since teacher-based assessment scores contribute to the calculation of USE-EE scores, they bring about affective consequences. Along with high-stakes testing policies, socio-economic dynamics like economic wealth and social approval affect the stakeholders' perceptions and expectations. Therefore, they expect EFL teachers to administer more exams and teaching to tests. Some institutional policies in MoNE influence institutional dynamics in lower secondary education. More particularly, school prestige and student recruitment are two basic outcomes and triggers of USE-EE. So, school administration expects EFL teachers to run supplementary courses and prepare the 8th graders socially, academically and psychologically for USE-EE and perform exam-oriented teaching in the 8th grade. In addition, EFL teachers become accountable for USE-EE results. How intense the washback effects operate at lower secondary schools also depends on teacher-related factors like professional identity, previous learning background and assessment literacy. Teachers have a tendency to teach and test as they were taught and tested. Also, they need to enhance their testing knowledge and skills to avoid only use of exam-oriented testing. Their identity of language teaching mediates the washback effects of USE-EE in the classroom. Finally, exam-oriented teaching not only stems from the

implementation of USE-EE but also is triggered by contextual factors (duration of teaching) shaped by curricular policies.

In upper secondary education, there is a more general framework of washback (see Figure 22). As a milestone of the framework, foreign language policies in Turkey are expected to determine both high-stakes and low-stakes testing policies in terms of test coverage. Then high-stakes assessment policies including the administration of FLT and BCT influence some other dynamics: low-stakes assessment policies and then large-scale test blue-prints and stakeholders' perceptions and expectations. Socio-cultural dynamics and employability policies also shape the stakeholders' expectations and perceptions on English language teaching and testing. There are some results obtained from this relationship. The stakeholders expect exam-driven language teaching, testing and test practicing. EFL teachers implement their own teaching curriculum consisting of mock exams, strategy-based training and focusing on exam practice. Influence of low-stakes assessment on high-stakes test results pushes EFL teachers to behave leniently especially in non-English major classes.

On the other hand, inclusion of skill-based assessment in low-stakes assessment promotes attainment of some language skills and triggers collaboration among EFL teachers. However, language learners are exposed to a methodological shift in the 10th and 11th grade because of the BCT coverage and FLT specifications. Similar to lower secondary education, institutional policies that affect institutional dynamics like school prestige and student recruitment enable the stakeholders to focus attention on the centralized exam results and develop some strategies to improve these results accordingly. Finally, teacher-related factors (i.e. assessment literacy, teacher beliefs, pre-service and in-service education) may trigger or mediate the washback effects of FLT. If they believe that language learners need to develop four skills, they bring communicative tasks into the classroom or assess these four skills communicatively. Otherwise, they cannot show resistance to the negative influence of high-stakes exams.

Implications

This part provides a list of implications and implementations which will improve English language teaching and make contributions to language testing. Also, further insights are given in line with the research findings.

- The requirement of four-skills assessment should be placed into examination regulations, and a guide including sample forms of language assessment and rubrics should be provided to EFL teachers.
- Four skill-assessment should be conducted especially in language courses, and the awareness on skill-based assessment should be promoted in the educational society.
- Classroom teachers and other stakeholders should be informed about the principles of communicative teaching and testing.
- In-service training about language assessment knowledge and skills should be provided to EFL teachers to improve their assessment literacy.
- Mock exams should have construct validity and include assessment tasks that are similar to teaching and learning activities.
- Pre-service education programs should carry out more practices to shape EFL teachers' teaching and testing missions, beliefs and identity.
- Teachers should not be only individuals that are accountable for test results; all the stakeholders should share the same responsibility.
- Power of centralized exams should be lessened to prevent language teaching and assessment from negative washback effects.
- In order to integrate skills-based assessment with large-scale testing, the current research model's LASBLA should be put into practice in order to meet the basic requirements of four skills assessment.
- The alignment of high-stakes test FLT with language teaching policies and curricula should be constructed.

REFERENCES

- Akpınar, K. D., & Cakildere, 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, J. C., & Wall, D. (1993). Does washback exist?. *Applied linguistics*, 14(2), 115-129.
- Ali, M. M., & Hamid, M. O. (2020). Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh?. *Language Assessment Quarterly*, 17(2), 129-146.
- Alptekin, C., & Tatar, S. (2011). Research on foreign language teaching and learning in Turkey (2005–2009). *Language Teaching*, 44(3), 328-353.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—a case-study. *System*, 30(2), 207-223.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 37-50). New Jersey: Lawrence Erlbaum.
- Arslan, R. S., & Üçok-Atasoy, M. (2020). An investigation into EFL teachers' assessment of young learners of English: Does practice match the policy? *International Online Journal of Education and Teaching*, 7(2), 468-484.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University.

- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, 2(1), 1-34.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language testing*, 13(3), 257-279.
- Bazeley, P., & Jackson, K. (2019). *Qualitative data analysis with NVivo: 2nd Edition*. Sage.
- Bloor, M., & Wood, F. (2006). *Keywords in qualitative methods: A vocabulary of research concepts*. Sage.
- Bogden, C. R., & Biklen, S. K. (2007). *Qualitative Research for Education: An Introduction to Theory and Methods*. Pearson.
- Broad, B. (2007). The power of tests: A critical perspective on the uses of language texts by Elana Shohamy. *Journal of Writing Assessment*, 3(1), 55-60.
- Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. Boston: Heinle & Heinle.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL quarterly*, 32(4), 653-675.
- Brown, J. D., Hudson, T., & Norris, J. Bonk, WJ (2002). *An investigation of second language task-based performance assessments*. Second Language Teaching & Curriculum Center.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York: McGraw–Hill ESL/ELT.
- Büyükkarıcı, K. (2016). Identifying the areas for English language teacher development: A study of assessment literacy. *Pegem Journal of Education and Instruction*, 6(3), 333-346.
- Carr, N. (2011). *Designing and analyzing language tests*. Oxford: Oxford University.
- Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher and F. Davidson (Eds.). *The Routledge handbook of language testing*, (pp. 65-78). New York: Routledge.

- Chen, L. M. (2002). Washback of a public exam on English teaching. *ERIC*, 1-21.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15(3), 253-271.
- Cheng, L. (2003). Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong case study. *The Journal of Classroom Interaction*, 1-10.
- Cheng, L., & Curtis, A (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanable, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-17). New Jersey: Lawrence Erlbaum.
- Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436-470.
- Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39-62.
- Cohen, L., & Manion, L. (2007). *Research methods in education (6th Ed.)*. New York: Taylor & Francis e-Library.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education (Sixth edition)*. Routledge.
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage.
- Creswell, J. W. (2014). *Educational research. Planning, conducting, and evaluating quantitative and qualitative research: 4th Edition*. Sage.
- Çakır, A. (2017). Türkiye’de YÖK öncesi ve sonrası yabancı dil eğitimi. *Journal of Language Research (JLR)*, 1(1), 1-18.
- Çakır, İ. (2017). The washback effects of secondary education placement examination on teachers, school administrators and parents with specific reference to teaching English as a foreign language. *Turkish Journal of Teacher Education*, 6(2), 61-73.

- Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics*, 29, 90-100.
- Davies, Alan, Annie Brown, Cathie Elder, Kathryn Hill, Tom Lumley, and Tim McNamara. (1999). *Dictionary of language testing. Vol. 7*. Cambridge: Cambridge University.
- Douglas, D. (2010). *Understanding language testing*. Hodder Education.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. New York: Oxford University.
- Ersöz, A., Çakır, A., Cephe, P. T., Peker, B. G., Özkan, N., Büge, C., & Özmen, D. (2006). *English language curriculum for primary education grades 4, 5, 6, 7, and 8*. Ankara: Talim ve Terbiye Kurulu Başkanlığı.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education (8th Ed.)*. New York: Mc Graw Hill.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Fulcher, G. (2009). Test use and political philosophy. *Annual Review of Applied Linguistics*, 29, 3-20.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.
- Genç, A. (2004). Türkiye’de ilk ve ortaöğretim okullarında yabancı dil öğretimi. *Kırgızistan-Manas Sosyal Bilimler Dergisi*, 10, 107-111.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University.
- Gillham, B. (2000). *Case study research methods: Real world research*. London: Paston Prepress Ltd.
- Göktürk-Sağlam, A. L. (2018). Can exams change how and what teachers teach? investigating the washback effect of a university English language proficiency test in the Turkish context. *Eurasian Journal of Applied Linguistics*. 4(2), 155-176.

- Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11(2), 113-134.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97.
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51.
- Güven, İ. (2018). *Türk Eğitim Tarihi (7. Baskı)*. Ankara: Pegem.
- Hatipoğlu, Ç. (2015). English language testing and evaluation (ELTE) training in Turkey: Expectations and needs of pre-service English language teachers. *ELT Research Journal*, 4(2), 111-128.
- Hatipoğlu, Ç. (2016). The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia-Social and Behavioral Sciences*, 232, 136-144.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University.
- Hung, S. T. A. (2012). A washback study on e-portfolio assessment in an English as a Foreign Language teacher preparation program. *Computer Assisted Language Learning*, 25(1), 21-36.
- İlköğretim ve Eğitim Kanunu, [Law of Education and National Education], *T.C. Resmi Gazete*, 10705, 12 January 1961.
- Kilickaya, F. (2016). Washback effects of a high-stakes exam on lower secondary school English teachers' practices in the classroom. *Lublin Studies in Modern Languages and Literature*, 40(1), 116.
- Kirkgoz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations. *RELC journal*, 38(2), 216-228.
- Kırkgöz, Y., Celik, S., & Arıkan, A. (2016). Laying the theoretical and practical foundations for a new elementary English curriculum in Turkey: a procedural analysis. *Kastamonu Eğitim Dergisi* 24(3), 1199-1212.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage.

- Külekçi, E. (2016). A concise analysis of the Foreign Language Examination (YDS) in Turkey and its possible washback effects. *International Online Journal of Education and Teaching*, 303-315.
- Luxia, Q. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education*, 14(1), 51-74.
- MacKey, A., & Gass, S. (2005). *Second language research: Methodology and design*. New Jersey: Lawrence Erlbaum Associates.
- Madsen, H. S. (1983). *Techniques in testing*. Oxford: Oxford University.
- McKay, S. L. (2006). *Researching second language classrooms*. London: Lawrence Erlbaum Associates.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University.
- Menken, K. (2017). High-Stakes tests as de facto language education policies. In E. Shohamy, I. G. Or, and S. May (Eds.), *Encyclopedia of language and education*, (pp. 385-396). Boston: Springer.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Millî Eğitim Bakanlığı Merkezî Sistem Sınav Yönergesi, [Ministry of National Education Regulation for Centralized Examination], 8649147, 12 August 2016.
- Millî Eğitim Bakanlığı Okul Öncesi Eğitim ve İlköğretim Kurumları Yönetmeliği. [Ministry of National Education Regulation for Pre-school and Primary Education Institutions], T.C. Resmi Gazete, 29072, 26 July 2014.
- Millî Eğitim Bakanlığı Ortaöğretime Geçiş Yönergesi. [Ministry of National Education Regulation for Transition to Upper Secondary Education], T.C. Resmi Gazete, 6738897, 3 April 2019.

- Millî Eğitim Bakanlığı Orta Öğretim Kurumları Yönetmeliği. [Ministry of National Education Regulation for Upper Secondary Education], *T.C. Resmi Gazete*, 28758, 07 September 2013.
- Millî Eğitim Bakanlığı Yabancı Dil Eğitimi ve Öğretimi Yönetmeliği. [Ministry of National Education Regulation for Foreign Language Education and Teaching], *T.C. Resmi Gazete*, 26184, 31 May 2006.
- Millî Eğitim Temel Kanunu. [Basic Law of National Education], *T.C. Resmi Gazete*, 14574, 24 June 1973.
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language testing*, 27(1), 33-49.
- Norton, B. (1997). Accountability in language assessment. In C. Clapham and D. Corson (Eds.), *Encyclopedia of language and education*, Vol. 7, (pp. 313-22). The Netherlands: Kluwer Academic.
- Ölçme, Seçme ve Yerleştirme Merkezi Hizmetleri Hakkında Kanun. [Law of Measurement, Selection and Placement Services (MSCP)]. *T.C. Resmi Gazete*, 27683, 3 March 2011.
- Ölçme, Seçme ve Yerleştirme Merkezi. [Measurement, Selection and Placement Center (MSCP)]. 2021 Yükseköğretim Kurumları Sınavı (YKS) Kılavuzu. [2021 Guide for Higher Education Examination].
- Önal, O., & Gürsoy, E. (2020). Primary and secondary level EFL teachers' use of assessment and assessment results in Turkey. *Journal on English Language Teaching (JELT)*, 10(3), 1-11.
- Özmen, K. S. (2011a). Analyzing washback effect of SEPPPO on prospective English Teachers. *Journal of Language and Linguistic Studies*, 7(2), 24-52.
- Özmen, K. S. (2011b). Washback effects 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.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Sage.

- Polat, M. (2020). YDS ve YÖKDİL sınavlarının yabancı dil eğitimi üzerine etkileri. *Kastamonu Education Journal*, 28(5), 2188-2202.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *The Phi Delta Kappan*, 68(9), 679-682.
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100, 190-208
- Ross, S. J. (2011). The social and political tensions of language assessment. In E. Hinkel (Ed.) *Handbook of research in second language teaching and learning*, II, (pp. 786-797).
- Sarıgöz, İ. H., & Fişne, F. N. (2018). English language assessment and evaluation practices in the 4th grade classes at main stream schools. *Journal of Language and Linguistic Studies*, 14(3), 380-395.
- Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. Teachers College.
- Serpil, Ö. Z., & Derin, A. T. A. Y. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Shaw, S. T., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge: Cambridge University.
- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9(3), 15-22.
- Sherman, R. R., & Webb, R. B. (Eds.). (1988). *Qualitative research in education: Focus and methods* (Vol. 3). Psychology.
- Shih, C. M. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8(2), 188-206.
- Shih, C. M. (2010). The washback of the General English Proficiency Test on university policies: A Taiwan case study. *Language Assessment Quarterly*, 7(3), 234-254.

- Shohamy, E. (1993). The Power of tests: The impact of language tests on teaching and learning. *NFLC Occasional Papers*, 1-19.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331-45.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education*, 14(1), 117-130.
- Shohamy, E. (2017). Critical language testing. Shohamy, E., Or, I. G., & May, S. (Eds.). (2017). *Language testing and assessment*, (pp. 441-454). Springer.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Soruların Hazırlanması, Soru Havuzunun Oluşturulması ve Soruların Güvenliğinin Sağlanmasına İlişkin Yönetmelik. [Regulation for Item Preparation, Item Pooling and Ensuring Item Security], *T.C. Resmi Gazete*, 28620, 16 April 2013.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Stake, R. E. (2006). *Multiple case study analysis*. Guilford press.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Şenel, E., & Tütüniş, B. (2011). The washback effect of testing on students' learning in EFL writing classes. *Dil Dergisi*, 153, 43-52.
- T.C. Millî Eğitim Bakanlığı. [Ministry of National Education (MoNE)]. (2021). *Sınavla Öğrenci Alacak Ortaöğretim Kurumlarına İlişkin Merkezî Sınav Başvuru ve Uygulama Kılavuzu*. [Guide for Application and Administration of Centralized Examination for Upper Secondary Education].
- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2013). Secondary education regulations. *T.C. Resmi Gazete*, 28758, 7 September 2013.
- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2013). *İlköğretim Kurumları (İlkokullar ve Ortaokullar) İngilizce Dersi (2, 3, 4, 5, 6, 7 ve 8. Sınıflar) Öğretim Programı* [English Language Teaching Program for primary and secondary schools (The 2nd, 3rd, 4th, 5th, 6th, 7th and 8th grade)]. Ankara: MEB.

- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2014). Pre-school, primary and secondary school regulations. *T.C. Resmi Gazete*, 29072, 26 July 2014.
- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2018). *İngilizce Dersi Öğretim Programı (İlkokul ve Ortaokul 2, 3, 4, 5, 6, 7 ve 8. sınıflar)*. [English language teaching program (Primary and Secondary schools the 2nd, 3rd, 4th, 5th, 6th, 7th and 8th grade)]. Ankara: MEB.
- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2018). *Ortaöğretim İngilizce Dersi (9, 10, 11 ve 12. Sınıflar) Öğretim Programı*. [English Language Teaching Program for secondary schools (The 9th, 10th, 11th and 12th grade)]. Ankara: MEB.
- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2021). *Milli Eğitim İstatistikleri: Örgün Eğitim 2020/21* (National Education Statistics: Formal Education 2020/21). Ankara: MEB.
- T.C. Milli Eğitim Bakanlığı [Ministry of National Education (MoNE)]. (2020). *Ortaöğretime geçiş ve yerleştirme kılavuzu*. [Booklet for selection and placement into upper secondary education)]. Ankara: MEB.
- Talim ve Terbiye Kurulu Başkanlığı. [Council of Education and Moral]. (2005). Orta öğretimin yeniden yapılandırılması. *Milli Eğitim Bakanlığı Tebliğler Dergisi*, 2573/184, 381-383.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36.
- Toksoz, I., & Kilickaya, F. (2017). Review of Journal Articles on Washback in Language Testing in Turkey (2010-2017). *Lublin Studies in Modern Languages and Literature*, 41(2).
- Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected Papers on Theoretical and Applied Linguistics*, 19, 431-445.

- Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I. G. Or, and S. May (Eds.), *Encyclopedia of language and education*, (pp. 359-372). Boston: Springer.
- Turner, C. E. (2012). Classroom Assessment. In G. Fulcher and F. Davidson (Eds.), *The Routledge handbook of language testing*, (pp. 65-78). New York: Routledge.
- Uzun, A., & Kiliçkaya, F. (2020). TEOG (TEPSE) English test: Content validity and teachers' views. *Bartın University Journal of Faculty of Education*, 9(3), 680-708.
- Van Ha, X., Tran, N. G., & Tran, N. H. (2021). Teachers' Beliefs and Practices Regarding Assessment in English as a Foreign Language Classrooms in Vietnam. *Qualitative Report*, 26(11).
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- Wang, J. (2011). *A study of the role of the teacher factor in washback*. Unpublished doctoral dissertation: McGill University.
- Watanable, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanable, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19-36). New Jersey: Lawrence Erlbaum.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.
- Wicking, P. (2017). The assessment beliefs and practices of English teachers in Japanese Universities. *JLTA Journal*, 20, 76-89.
- Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly* 45(4), 628-660.
- Wiyaka, W. (2020). Washback of the performance-based test of reading for EFL learners. *EduLite: Journal of English Education, Literature and Culture*, 5(2), 202-213.
- Xiao, Y., Sharpling, G., & Liu, H. (2011). Washback of national matriculation English test on students' learning in the Chinese secondary school context. *Asian EFL Journal*, 13(3), 103-129.

- Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, 50, 56-68.
- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49-70.
- Yabancı Dil Eğitimi ve Öğretimi ile Türk Vatandaşlarının Farklı Dil ve Lehçelerinin Öğrenilmesi Hakkında Kanun. [Law of Foreign Language Teaching and Turkish Citizens' Learning of Different Languages and Dialects] *T.C. Resmi Gazete*, 18196, 19 October 1983.
- Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but...": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168.
- Yıldırım, A., & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin Yayıncılık.
- Yildirim, O. (2010). Washback effects of a high-stakes university entrance exam: Effects of the English section of the university entrance exam on future English language teachers in Turkey. *The Asian EFL Journal Quarterly*, 12(2), 92-116.
- Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage publications.
- Yücel, E., Dimici, K., Yıldız, B., & Bümen, N. (2017). Son 15 yılda yayımlanan ilk ve ortaöğretim İngilizce dersi öğretim programları üzerine bir analiz. *Ege Eğitim Dergisi*, 18(2), 702-737.
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71-89.
- Zou, S., & Xu, Q. (2017). A washback study of the test for English majors for grade eight (TEM8) in China—From the perspective of university program administrators. *Language Assessment Quarterly*, 14(2), 140-159.



APPENDICES

APPENDICES

Appendix A

“Washback Effect of Language Testing on EFL Teaching”

Interview Form

Date:/...../.....

Hour:

Research Questions

- 1) What are the impacts of language testing on macro- and micro-level dynamics of EFL teaching?
- 2) What are the triggers that bring about the impacts of language testing?

INTRODUCTION

Dear Participant,

I am Fatıma Nur Fişne and work at English Language Teaching Program, Gazi Faculty of Education, Gazi University. I am conducting a research study about ***the impacts of language testing on EFL teaching and the reasons for these impacts***. In this regard, I would like to interview with English language teachers who work at the state schools in the Ministry of National Education (primary, secondary and high schools). There are two types of language testing that this research focuses on: ***teacher-based language assessment and large-scale language testing (high-stakes language testing)***. Confidentiality will be ensured throughout the interview. If you give permission, I am going to record this interview. This record will not be definitely used for any other reason except for the research aims. Your personal data will not be given in the research report. Instead, different codes will be used. You have the right to stop the record whenever you like and delete the part you do not want to share. Interview will take approximately 60-90 minutes.

Thank you very much for your participation and valuable time.

PART 1

Demographics

School Type:	Primary school ☐ Secondary school ☐ High school ☐	Graduation:	ELT ☐ Linguistics ☐ Literature ☐ Others ☐
Age:	_____	Years of teaching:	_____ years
Gender:	_____	Which grades do you teach?	_____
City/Province:	_____	Have you ever taken any training on testing?	Yes ☐ No ☐

PART II

SEMI-STRUCTURED INTERVIEW (EFL Teachers)*

1. Could you please give information about assessment and evaluation methods you administer in your classes? **
 - ✓ What kinds of methods do you use in different grades?
 - ✓ What are your assessment and evaluation practices in different grades?
 - ✓ What are the basic materials of the assessment and evaluation in your classes?
2. How consistent are your assessment practices consistent with language objectives?
 - ✓ Speaking objectives
 - ✓ Listening objectives
 - ✓ Writing objectives
 - ✓ Reading objectives
3. How consistent are your assessment practices consistent with teaching methodology?
 - ✓ Course content
 - ✓ Teaching practices
 - ✓ Language teaching methods
 - ✓ Language materials
4. Do you think that your assessment practices influence your teaching methodology? If yes, could you please elaborate on this influence?
 - ✓ Course content
 - ✓ Teaching practices
 - ✓ Language teaching methods
 - ✓ Language materials
5. How important are the classroom-based assessment practices for the stakeholders?
Alternative question: What is the importance that the stakeholders give to the assessment procedure?
 - ✓ For you
 - ✓ For your students
 - ✓ For parents
 - ✓ For the school administration
 - ✓ For your colleagues
 - ✓ Why?
 - ✓ Could you please explain it with examples?
6. Do the stakeholders have any expectations from the classroom-based assessment practices? If yes, what are they?
Alternative question: How assessment procedure do the stakeholders expect you to follow?
 - ✓ Your students
 - ✓ Parents
 - ✓ School administration
 - ✓ Your colleagues

- ✓ Why?
 - ✓ Are your assessment practices influenced by these expectations?
 - ✓ Are your teaching practices influenced by these expectations?
7. Do you and your colleagues work in cooperation to develop and then evaluate assessment practices? If yes, could you please give more information on this cooperation?
- ✓ How is this cooperation?
 - ✓ How often do they cooperate with each other?
8. Do you and your institution make any decisions in consideration of the assessment results? If yes, what are they?
- ✓ What kinds of decisions do you make?
 - ✓ Who is affected by these decisions?
9. To what extent does large-scale centralized examination influence the attainment of language objectives?
- Alternative question: Does large-scale centralized examination facilitate the attainment of language objectives?
- ✓ Speaking objectives
 - ✓ Listening objectives
 - ✓ Writing objectives
 - ✓ Reading objectives
10. Does large-scale centralized examination influence your teaching methodology? If yes, could you please give information how it influences?
- ✓ Course content
 - ✓ Teaching practices
 - ✓ Language teaching methods
 - ✓ Language materials
11. Does large-scale centralized examination influence your assessment and evaluation procedure? If yes, could you please give information how it influences?
- ✓ Assessment and evaluation methods
 - ✓ Assessment and evaluation practices
 - ✓ Assessment and evaluation materials
12. How important is large-scale centralized examination for the stakeholders?
- Alternative question: What is the importance that the stakeholders give to the large-scale centralized examination?
- ✓ For you
 - ✓ For your students
 - ✓ For parents
 - ✓ For the school administration
 - ✓ For your colleagues
- ✓ Why?

- ✓ Could you please explain it with examples?

13. What are the stakeholders' expectations from English language teaching in consideration of the large-scale centralized examination?

- ✓ Your students
- ✓ Parents
- ✓ School administration
- ✓ Your colleagues

- ✓ Course content
- ✓ Teaching practices
- ✓ Language materials
- ✓ Assessment and evaluation practices
- ✓ Is your assessment procedure influenced by these expectations?
- ✓ Is your teaching procedure influenced by these expectations?

14. Do you and your colleagues work in cooperation in consideration of the large-scale centralized examination? If yes, could you please give information on this cooperation?

- ✓ How is this cooperation?
- ✓ How often do they cooperate with each other?

15. Does your institution make any decisions in consideration of the large-scale centralized examination results? If yes, could you please elaborate on these decisions?

- ✓ What kinds of decisions does your institution make?
- ✓ Who is affected by these decisions?

*Interview questions 1 to 8 will be directed to EFL teachers who work at primary school.

**** Basic Terminology of Assessment and Evaluation**

Assessment: It is a process of collecting information about target performance to define the level of language attainment.

Evaluation: It is a decision-making process by comparing the assessment results with a certain criterion or criteria set.

Assessment methods: Self-assessment, portfolio-assessment, traditional exams, projects, task-based language assessment, conferences, observation, performance-based assessment etc.

Assessment practices: Language tasks, project assignments, checklists, written exams, quizzes, performance tasks, games etc.

“Washback Effect of Language Testing on EFL Teaching”

Interview Form

Date:/...../.....

Hour:

Research Question

- 1) What are the triggers that bring about the impacts of language testing on macro- and micro-level dynamics of EFL teaching?

INTRODUCTION

Dear Professor,

I am Fatıma Nur Fişne and work at English Language Teaching Program, Gazi Faculty of Education, Gazi University. My research study is about ***the impacts of language testing on EFL teaching***. There are two types of language testing that this research focuses on: ***teacher-based language assessment and large-scale centralized language testing***. Confidentiality will be ensured throughout the interview. If you give permission, I am going to record this interview. This record will not be definitely used for any other reasons except for the research aims. Your personal data will not be given in the research report. Instead, different codes will be used. You have the right to make the record stop whenever you like and delete the part you do not want to share. Interview will take approximately 60-90 minutes.

Thank you very much for your participation and valuable time.

PART 1

Demographics

University:	_____	City:	_____
Age:	_____	Gender:	_____
Work experience:	_____ years	Do you work as an academic counsellor in any state school?	Yes ☐ No ☐
Graduation (BA, MA, PhD)	ELT ☐ Linguistics ☐ Literature ☐ Others ☐	Have you ever worked as an EFL teacher at any state school in the MoNE?	Yes ☐ No ☐
Do you train English language teachers on assessment and evaluation?	Yes ☐ No ☐	Do you train student teachers on English language assessment and evaluation?	Yes ☐ No ☐

PART II

SEMI-STRUCTURED INTERVIEW (EFL Teacher Trainers/Field Experts)

1. Centralized exams measure only reading comprehension, grammar and lexical use. However, language teaching curriculum puts emphasis on assessment and evaluation of all language skills and components.
 - ✓ Why do not centralized exams measure all language skills and components?
 - ✓ Why do you think that there is a discrepancy between testing and teaching policies (curriculum)?
 - ✓ Is there any language testing policy in Turkey? If yes, what is it?
 - ✓ What influences our centralized language testing policy?
2. In the regulation of assessment and evaluation in primary and lower secondary education (Grade 4 to 8), there is an emphasis on the number and administration of written exams. In upper secondary education (Grade 9 to 12), the regulation requires performance-based assessment in language classes in addition to written exams. These regulations influence classroom-based language testing.
 - ✓ Do you think that language testing policy is employed in the development of these regulations?
 - ✓ What factors are taken into consideration in the development of assessment and evaluation regulation in each level of education?
3. EFL teachers' mission, assessment literacy, beliefs on language teaching and testing influence how they teach English and assess and evaluate their students in the classroom.
 - ✓ What shapes teacher-related factors in language testing?
4. In society, there are some perceptions on language teaching and testing. To take an example, teachers should not spend much time on listening to songs and playing games. Instead, teachers should present language items, administer language tests and give more homework. Language teachers are expected to use more traditional methods in teaching and testing.
 - ✓ How are these perceptions shaped in the society?
 - ✓ Is it possible to change these perceptions? If yes, how?
5. Language exams, particularly centralized exams, have power in society.

For students, they pave the way for getting a job and earning money, gaining respect in the society, making dreams come true. For language teachers, they reflect course achievement. For school administration, they give evidence on teaching quality.

 - ✓ What are social, psychological, economic and cultural consequences of these centralized language exams?
 - ✓ For students, teachers, parents and community
 - ✓ Do you think that centralized language exams have any impact on language teaching policies?
6. Do you want to add any other comments or information about the washback effects of language testing?
 - ✓ If yes, what are they?

Appendix B

EXPERT OPINION FORM

Dear,

This form has been prepared to measure the content validity index/ratio of semi-structured interview questions about “*Washback Effects of Language Testing on the Dynamics of EFL Teaching*”. In this study, it is aimed to investigate the macro- and micro-level dimensions of the effects of language testing on EFL teaching, stakeholders (students, parents, school administration, policy makers), language education system and policies. In order to describe the influence of language testing (Interview Form 1), interview will be conducted with EFL teachers who work at primary, secondary and high school in MoNE. After that, the reasons for these effects will be discussed with field experts (Interview Form 2). There are two main factors of this study: *teacher-based language assessment* and *large-scale centralized language testing*. Interview 1 will be carried out in Turkish language to avoid language-related concerns. That’s why Turkish translation is given in brackets.

You are kindly requested to examine the interview questions and put a tick on the relevant box in consideration of the appropriateness of the interview questions to the content of the research. If you think that a question should be improved, please explain the points that need improvement in this question.

Thank you for your valuable contribution.

Res. Assist. Fatıma Nur FİŞNE

INTERVIEW FORM 1 (EFL Teachers)

1) Inappropriate 2) Appropriate but needs improvement 3) Appropriate

	(1)	(2)	(3)

Appendix C

Category: Methodological Shift in Language Assessment (Case 1)

ENGLISH	TURKISH
There are some differences between primary and lower secondary education. We are not expected to administer written exams [in the 2nd and 3rd grades] at primary schools but there should be written documents of the exams at lower secondary schools. At primary schools, as we [practiced] at university, we employ some tasks and get them to prepare learning portfolios. In this way, they can see what they have learnt. Besides, we carry out performance-based assessment. We use drama, role-plays and songs that are important for us in assessment and evaluation. ... We do not prefer using tests in written exams [in the 4th grade]. They are based on matching and comprehension. [PLST 2]	İlkokul ve ortaokul çok çok farklı bir mecra. Bizden ilkokulda ölçme aracı olarak yazılı kağıt istemiyorlar ama ortaokulda bir yazılı kağıdı bir evrak olması lazım, ilkokulda biraz daha üniversitede öğrendiğimiz tarzda kendilerinin gelişim dosyalarının olması... işte etkinlik yapıyor öğrenciler daha sonra kendilerinin neler yaptıklarını görebiliyorlar, daha sonra performans değerlendirme yapıyoruz, tahtanın önüne çıkıp rol oynama drama yapma tarzında böyle daha çok etkinlik yapma fırsatımız oluyor, şarkı dinleme bunlar daha çok önemli ölçmede öğrencileri değerlendirme de bizim için daha çok önemli oluyor... Yazılılarda da genelde test kullanmamaya özen gösteriyoruz hani çocuklar matching yapsın, paragrafları nasıl çözebileceğini öğretsin. [PLST 2]

Category: Exam-Oriented Language Teaching (Case 1)

ENGLISH	TURKISH
While we enjoy language learning in the 2nd and 3rd grades, we feel like in the 4th grade that these students will take exams and we cannot test what we do not teach in the 4th grade. For example, we are going to administer two exams in the fall term; so, we have to cover the first 4-5 units as they will take two exams in the fall term. In this way, I teach what I plan to test. Thus, I give more weight to the exams and teach in line with them. [PST 2]	2. ve 3. Sınıfta normal giderken çok güzel eğlenerek öğrenirken 4. Sınıfta artık öğretmen olarak bizlere de şöyle bir hissiyat geçiyor yani bu çocuklar sınav olacaklar, ben onlara vermediğim bir şeyden sınav yapamam, örnek veriyorum ilk 4 ya da 5 ünite, ilk dönem iki sınav olacaklarsa bunların hepsini yapmak zorundayız ki sınavda soracaklarımı ben onlara sorabileyim Bundan dolayı biz de biraz sınava ağırlık veriyoruz, sınava yönelik çalışmalar yapmaya başlıyoruz. [PST 2]

Category: Parental Expectations (Case 1)

ENGLISH	TURKISH
Some parents think that these scores serve as the basis for their children's further success. Some others, let's say who have personal ambitions, think that their children are very successful, and they do not accept the failure [the score below a specific level]. They think that their children must be always successful. I believe that there are two groups, one of which is the parents who have biggest personal ambitions, and other one of which is the parents who think that the scores are important for the following years since they could be a learning habit. [PST 3]	Bazılarına puanların, çocuğun diğer yıllardaki başarılarına temel teşkil edeceğini düşündüğü için olanlar var, bir de bence bir grupta şahsi hırsları diyeyim, çocuğum çok başarılı deyip bunda altında olmasını kabul etmeyenler veli profili vardır ya benim çocuğum hep iyi olmalı diyen veliler var. Bence bu iki grup oluyor genelde biri şahsi başarı tutkusu yüksek olan veliler bir diğeri bunun çocuğu için bir alışkanlık olacağını diğer yıllarda temeli teşkil ederken bunların bir önemi olduğunu düşünen grup. [PST 3]

Category: Washback Effects of Practice Tests and Triggering Factors (Case 1)

ENGLISH	TURKISH
<i>I have witnessed something this week that one of the parents would like to change her child's [language] teacher because this teacher brings games to the classroom, the students have fun and laugh. She hesitates about whether the teacher gives lectures or not. The parents' perceptions on learning are that teachers employ tests, want the students to write something, assign them [some] homework etc. For this reason, she wants her child to be trained by another teacher. They think that he/she is not a good teacher because of using songs, tasks or draw-and-paint activities. [PST 1]</i>	<i>Bu hafta bir velinin söylemine şahit oldum, çocuğunun artık o öğretmende olmasını istemiyormuş, gerekçe olarak da öğretmen çok oyun oynatıyormuş, çocuğu çok eğleniyormuş, gülüyormuş... Acaba ders işlenmiyor mu? Velinin bakış açısı, öğretmen dediğin test çözdürür, yazı yazdırır, ödev yollar... O yüzden çocuğun artık o öğretmende olmasını istemiyormuş. Şarkı söylüyorlar, etkinlik yapıyorlar, çiziyorlar, boyuyorlar diye iyi bir öğretmen olmadığını düşünüyor. [PST 1]</i>

Category: Teacher-Based Assessment and Triggering Factors (Case 2)

ENGLISH	TURKISH
<i>The students' diploma score plays a significant role in transition to upper secondary education. For this reason, it is getting more important in the 8th grades. [LST-V-2]</i>	<i>Şöyle çocukların okul başarı puanları da önemli olduğu için tabi ki liseye geçiş sınavlarını da etkileyeceği için 8. Sınıflarda biraz daha önemli oluyor. [LST-V-2]</i>

Category: Methodological Shift in Language Teaching (Case 2)

ENGLISH	TURKISH
<i>There are some changes in the 8th grade. Our priority becomes the centralized exam. We focus on the exam vocabulary, item types, test-taking strategies in language teaching. On the other hand, I still use songs if they are related to the course content or I use word games in our revisions. [LST-2]</i>	<i>8'de şöyle bir şey değişiyor, sınav faktörü devreye girdiği için önceliğimiz sınav oluyor, sınavda çıkacak kelimeler sınavda çıkabilecek kalıplar sorular, test çözme tekniklerine odaklanıyorsunuz... ama tabi mesela şarkıları üniterim ile alakalı ise şarkı kullanıyorum hala mesela 8. Sınıflarda ders öğretirken tekrarlarımı yaparken oyunlar işin içine giriyor kelime oyunları yapıyoruz. [LST-2]</i>

Category: Teacher Professional Identity and Assessment Literacy (Case 2)

ENGLISH	TURKISH
<i>I have never assessed the listening skill in my formal exams. Some teachers assess it but I do not. [PLST-1]</i>	<i>Sınavlarımda dinleme de hiç sormadım, bazı soran arkadaşlarım var evet, sınavda dinleme partı da koyuyorlar ama ben listening hiç koymadım. [PLST-1]</i>

Category: Stakeholders' Perceptions and Expectations (Case 2)

ENGLISH	TURKISH
<i>The parents do not think games and songs are effective in teaching and they make contribution to the exams. They are partially right because the [centralized] exams do not measure the objectives that can be attained through these games and songs. [Field Expert 4]</i>	<i>Bir veli şarkıyla oyunlarla geçirilen zamanı çok verimli olduğunu düşünüyor, çünkü ona ne katkısı olacak deniyor, sınavların şekli de o tür aktivitelerden elde edilen bilgileri fazla ölçmediği için biraz haklılık payı da var. [Field Expert 4]</i>

Category: Washback Effects of USE-EE on Policy-Making (Case 2)

ENGLISH	TURKISH
<i>In our meetings [held] at the beginning of the academic year, we discuss what we [can] do [for the 8th graders]. We plan to run supplementary courses and conduct mock exams. In these courses, we completely do practices in line with the centralized exam. [LST-1]</i>	<i>Sene başında kurul toplantısında biz ne yapalım sorusu oluyor, bu çocuklara kurslar açalım, deneme günleri yapalım ve bu kurslarda da tamamen sınava yönelik çalışma yapmak. [LST-1]</i>

Category: Teacher Behaviours (Case 2)

ENGLISH	TURKISH
<i>There is a fear of not covering all the units given in the curriculum. What happens if it [the unit or topic not covered in the lesson] is tested in the centralized examination? Due to this [feeling of] fear, we have concerns about covering all the curricular content. [LST-3]</i>	<i>Müfredatı uyguladım mantığı bizim için bir korku, ne yazık ki, acaba müfredat yetişmezse korkusu. Ya bu bunun karşısına sınavda çıkarsa, korkusuyla o müfredatı yüzde yüz işleme kaygımız var. [LST-3]</i>

Category: Feedback on Language Learning (Case 3)

ENGLISH	TURKISH
<i>I have been working at this school since last year. When I met the students, I realized that the students participated in the lesson for the sake of being scored. When they did not attend the activities, I asked them the reasons why they did not take participation. They answered that they were not good at English lesson. I have deduced from this answer that they make judgments on whether their level of English [language] is 'good' or 'bad'" [on the basis of exam scores]. [UST-A-WE-2]</i>	<i>Ben size söylemiş olduğum kurumda geçen seneden beri çalışıyorum, öğrenciler ile ilk karşılaştığım zaman şunu farkettim ders içerisinde öğrencilerin daha çok derse katılma sebebi not üzerineydi. Diyorum ki derslere niye katılmıyorsun, Hocam ben İngilizce de iyi değilim ki dedi. Yani ben buradan şunu çıkardım, öğrenciler sanki almış olduğu notlar neticesinde İngilizcesinin iyi veya kötü olduğunu düşünüyor. [UST-A-WE-2]</i>

Category: High-Stakes and Low-Stakes Assessment Policies (Case 3)

ENGLISH	TURKISH
<i>You know that there is diploma score in our system. The higher the diploma score is, the better university you are placed in. Since they [students] have this awareness, it is very normal that the students have exam anxiety. [UST-S-E-2]</i>	<i>Sistemimizin [gerekliliği] ortalama var biliyorsunuz, ortalama ne kadar iyiye o kadar iyi bir üniversiteye girme şansı [var]. Bunu bildikleri için tabi ki çocukların not kaygılarının olması da çok normal. [UST-S-E-2]</i>

Category: Methodological Shift in Language Teaching (Case 3)

ENGLISH	TURKISH
<i>There is a sharp contrast here. This is not what I plan to teach. I want the students to [communicate in the target language] because I think that language is the life itself, not a lesson. [However] in language classes, we</i>	<i>Çok büyük bir tezat var burada, benim yapmak istediğim şey bu değil ki. Benim yapmak istediğim çocuklar ile dili [kullanma] peşindeyim çünkü dil hayatın kendisi dil bir ders değil benim için... Dil sınıflarında</i>

teach what is to be tested in FLT. To be honest, I am not happy with it but university entrance exam drives us to teach in this way. When I teach listening and speaking here, students will make objections [on these practices] because what I have taught is not measured in FLT. They are right. There are so many grammar topics to learn but not much time. They also have to take so many language tests. [UST-A-WE-3]	çocuklara sadece sınav onlara ne gerekiyorsa onları anlatıyoruz, bu durumdan ben de memnun değilim açıkça konuşayım ama sizin önünüzdeki üniversite sınavı maalesef bizi bunu yapmaya itiyor, kalkıp ben çocuklara burada listening ya da speaking yapsam çocuk bana kalkıp diyecek ki Hocam sen bana bunları anlattın, ama sınavda hiç böyle bir şey çıkmadı, haklı olarak yeteri zaman da yok çok fazla grammar konusu var, çok fazla test çözmeleri gerekiyor. [UST-A-WE-3]
--	---

Category: Socio-Cultural Dynamics and Employability Policies (Case 3)

ENGLISH	TURKISH
There is a university entrance exam. Everyone gets prepared for this exam with the purpose of getting a job and better life. That is why the exam itself is effective [in shaping the stakeholders' perspectives]. [Field Expert 6]	Üniversite sınavı gibi bir sınav var. Özellikle herkes tabi ki iş ve gelecek düşüncesi ile bu sınavlara hazırlanıyor, o yüzden de sınavın kendisi bence çok etkili bu alanda.. [Field Expert 6]

Category: High-Stakes Test Blueprints (Case 3)

ENGLISH	TURKISH
There should not be made any mistakes in the [centralized] exam that millions of people take. Just imagine what happens if there is a cut in electricity. The exam is cancelled. [In order to avoid such problems], you can take measures like use of batteries. [Field Expert 9]	[Milyonlarca] kişinin girdiği... bir sınav hatayı affetmiyor, düşünün dinleme sırasında bir elektrik gittiğinde. sınav iptal oldu... Önlem alınabilir: güç desteği, aküler. [Field Expert 9]
It will be successful when there is a match between what you teach in the school and what is measured in those exams. [Field Expert 7]	Okulda öğrettiğin ve girdiği sınavdaki sistem birbirleri ile örtüştüğü zaman zaten başarılı olacak. [Field Expert 7]

Category: Teacher Beliefs, Practices and Assessment Literacy (Case 3)

ENGLISH	TURKISH
I [observed] that the students [in the 12th grade – language classes] have weaknesses in speaking as an outcome of exam preparation. Thus I brought the songs and live news, and I tried to present authentic materials. We taught speaking and listening and exposed our students [to writing] by providing samples to construct background information. As for the examination, I assessed speaking skills and showed my rubrics to them. On the other side, I conducted written exams similar to the exam format [FLT] including cloze test, paragraph, irrelevant sentence at the level of grammatical, semantic or lexical level. [UST-A-WE-5]	Çocuklarımızın dinleme ve konuşmada son derece zayıf oldukları idi. Ben bunun tabi bir sınava hazırlanış biçiminin meyvesi olduğunu bildiğim için ... şarkılar buldum, canlı haber buldum yani authentic material sunmaya çalıştım onlara, ... Olaya bir speaking attık, listening attık [writing] sample larını sunup background oluşturması açısından exposure yaptık, sınavlara gelince de sözlülerimin mutlaka oldu konuşmaya rubriklerimi de gösterdim onlara. Diğer kısım ise gramatik, semantik, lexical hangi boyutta ise onları üniversiteye giriş sınavlarında maruz kalacakları reading gibi ... paragraf anlam bütünü bozmayan cümle gibi sınav formatı [kullandım]. [UST-A-WE-5]

Appendix D

Evrak Tarih ve Sayısı: 23.12.2020-E.137835



**T.C.
GAZİ ÜNİVERSİTESİ
Ölçme Değerlendirme Etik Alt Çalışma Grubu**

Sayı : E-91610558-302.08.01-
Konu : Bilimsel ve Eğitim Amaçlı

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 19.11.2020 tarihli ve E.124096 sayılı yazı.

İlgi yazınız ile göndermiş olduğunuz, Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı, İngilizce Dili Eğitimi Bilim Dalı **Doktora Öğrencisi Fatıma Nur FİŞNE'nin, Prof.Dr.Kemal Sinan ÖZMEN'in** danışmanlığında yürüttüğü **"Ölçme ve Değerlendirme Etkisi Türkiye'deki İngiliz Dili Eğitimi Makro ve Mikro Düzey Dinamikleri Üzerinde Nasıl Faaliyet Göstermektedir: İngilizce Öğretmenlerinin Ölçme ve Değerlendirme İnançları, Okur-Yazarlıkları ve Uygulamalarına Yönelik Görüşler"** adlı tez çalışması ile ilgili konu Kurulumuzun **01.12.2020** tarih ve **12** 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: 2020-689

Prof. Dr. İsmail KARAKAYA
Kurul Başkanı

Ek:1 Liste

Belge Doğrulama Kodu :BEACACSLC

Bu belge, güvenli elektronik imza ile imzalanmıştır.

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



Emniyet Mahallesi Bandırma Caddesi No :6/1 06560 Yenimahalle/ANKARA
Tel:0 (312) 202 20 57 - 0 (312) 2... Faks:0 (312) 202 38 76
İnternet Adresi :<http://etikkomisyon.gazi.edu.tr/>
Kep Adresi : gaziuniversitesi@ho01.kep.tr

Bilgi için :Esengül BOŞNAK
Birim Evmk Sorumlusu
Telefon No:03122022666

Bu belge,güvenli elektronik imza ile imzalanmıştır.

Evrak Tarih ve Sayısı: 18.05.2021-E.89090



T.C.
MİLLÎ EĞİTİM BAKANLIĞI
Strateji Geliştirme Başkanlığı

Sayı : E-49614598-605.01-25208975
Konu : Araştırma Uygulama İzni Talebi

Ek-1
17.05.2021

DAĞITIM YERLERİNE

- İlgi: a) Gazi Üniversitesi Rektörlüğünün 23/03/2021 tarihli ve E-80287700-302.08.01-58451 sayılı yazısı.
b) Millî Eğitim Bakanlığı'nın 21/01/2020 tarihli ve 2020/2 Nolu Araştırma Uygulama İzinleri Genelgesi.

İlgi (a) yazı ile Gazi Üniversitesi Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Ana Bilim Dalı Doktora öğrencisi Fatma Nur FİŞNE'nin "How Washback Operates in Macro- and Micro- level Dynamics of Efl Teaching in Turkey: Insights From Efl Teachers Testing Beliefs, Literacy and Practices (Ölçme ve Değerlendirme Etkisi Türkiye'deki İngiliz Dili Eğitimi Makro ve Mikro Düzey Dinamikleri Üzerinde Nasıl Faaliyet Göstermektedir: İngilizce Öğretmenlerinin Ölçme ve Değerlendirme İnançları, okur-yazarlıkları ve Uygulamalarına Yönelik Görüşler)" konulu araştırmasına veri sağlamak amacıyla anket çalışması yapma izin talebine ilişkin yazı ve ekleri Başkanlığımız tarafından incelenmiştir.

Araştırmanın virüsün etkilerinden korunmak amacı ile Bakanlığımıza bağlı resmi/özel okul ve kurumlarda öğrenci, öğretmen ve okul yöneticilerinin katılımı ile yapılması planlanan il/ilçe, ulusal ve uluslararası düzeydeki sosyal etkinliklerin (toplantı, çalıştay, sempozyum, konferans, forum, ödül töreni, spor müsabakası, yarışma vb.) tedbiren iptal edilmesi ve yüz yüze eğitim öğretime ara verilmesi göz önüne alınarak örgün eğitimin tam olarak başlamasıyla birlikte ilgili genel müdürlüklerin izni ile denetimi il/ilçe millî eğitim müdürlükleri ve okul/kurum idaresinde olmak üzere, kurum faaliyetlerini aksatmadan, gönüllülük esasına göre; onaylı bir örneği Bakanlığımızda muhafaza edilen ve uygulama sırasında da mühürlü ve imzalı örnekten çoğaltılan veri toplama araçlarının uygulanmasına ilgi (b) Genelge doğrultusunda izin verilmiştir.

Gereğini bilgilerinize rica ederim.

Mehmet Fatih LEBLEBİCİ
Bakan a.
Başkan

Ek: Onaylı Veri Toplama Araçları (12 Sayfa)

Dağıtım:

Gereği:

Adana, Adıyaman, Ankara, Artvin, Bartın, Bitlis, Bursa,
Burdur, Denizli, Erzurum, Gaziantep, İstanbul, Kayseri,
Kocaeli, Konya, Manisa, Mardin, Muğla, Samsun, Sivas,
Tokat, Van, Yozgat Valiliklerine (İl Millî Eğitim Müdürlüğü)

Bilgi:

Gazi Üniversitesi Rektörlüğüne

Bu belge güvenli elektronik imza ile imzalanmıştır.

Adres : MEB Atatürk Bulvarı No:98 Bakanlıklar/ANKARA

Belge Doğrulama Adresi : <https://www.turkiye.gov.tr/meb-dby5>

Telefon No : 0 (312) 413 27 53

E-Posta : sgb_araştırma_izinleri@meb.gov.tr

Keş Adresi : meb@hs01.kep.tr

Bilgi için: İstatistikçi Özlem BEYAZPİRİNÇ

İnternet Adresi: www.meb.gov.tr

Unvan : İstatistikçi

Faks: 312 418 64 01

Bu evrak güvenli elektronik imza ile imzalanmıştır. <https://evmlsorgu.meb.gov.tr> adresinden 53a1-5d51-33b2-9aed-d8fa kodu ile teyit edilebilir.

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Evrak Tarih ve Sayısı: 01.04.2021-E.64964



T.C.
PAMUKKALE ÜNİVERSİTESİ
Öğrenci İşleri Daire Başkanlığı

Sayı : E-93282220-302.08.01-38006
Konu : Uygulama İzni Hk (Fatma Nur FİŞNE)

GAZİ ÜNİVERSİTESİ REKTÖRLÜĞÜNE

- İlgi : a) Üniversiteniz Eğitim Bilimleri Enstitüsü Müdürlüğünün 24.03.2021 tarih ve E-80287700-302.08.01-58408 sayılı yazısı.
b) Üniversitemiz Eğitim Fakültesi Dekanlığının 29.03.2021 tarih ve E.37516 sayılı yazısı.

Üniversiteniz Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Anabilim Dalı Doktora Öğrencisi Fatma Nur FİŞNE'nin, "How Washback Operates in Macro-And Micro-level Dynamics of Efl Teaching in Turkey: Insights From Efl Teachers Esting Beliefs, Literacy and Practices" konulu tez çalışması kapsamında, Üniversitemiz Eğitim Fakültesi Yabancı Diller Eğitimi Bölümünde ölçek uygulama isteği Rektörlüğümüzce uygun görülmüştür.

Gereğini bilgilerinize arz ederim.

Prof. Dr. Necip ATAR
Rektör a.
Rektör Yardımcısı

Ek: İlgi (b) Yazı ve Ekleri (... sayfa)

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu :BSE67PRJL Pin Kodu :26092

Belge Takip Adresi : <https://www.turkiye.gov.tr/pau-ebys>

Adres:Kınıklı Yerleşkesi Rektörlük Binası 20160/DENİZLİ

Telefon:0 (258) 296 21 51 Faks:0 (258) 296 23 32

e-Posta:oidi@pau.edu.tr Elektronik Ağ:<http://www.pau.edu.tr/oidb>

Kep Adresi: pau.rektorkluk@bs01.kep.tr

Bilgi için: Necla GEBEŞ

Unvanı: Bilgisayar İşletmeni





GAZİLİ OLMAK AYRICALIKTIR...