REPUBLIC OF TURKEY

ÇAĞ UNIVERSITY

INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF ENGLISH LANGUAGE EDUCATION

THE IMPACT OF INSTITUTIONAL TESTS ON TEACHERS'

TEACHING METHODOLOGIES

THESIS BY

Songül Yeliz SARI

Supervisor- Head of examining committee: Prof. Dr. Şehnaz ŞAHİNKARAKAŞ

Member of examining committee: Dr. Seden TUYAN

Member of examining committee: Dr. Figen YILMAZ

(Çukurova University)

MASTER OF ARTS

MERSİN / 2018

REPUBLIC OF TURKEY

ÇAĞ UNIVERSITY

DIRECTORSHIP OF THE INSTITUTE OF SOCIAL SCIENCES

We **certify** that this thesis under the title of **"The Impact of Institutional Tests on Teachers' Teaching Methodologies"** prepared by **Songül Yeliz SARI** with number **20118016** is satisfactory for the award of the degree of **Master of Arts** in the Department of **English Language Education**.

....

(University-Inside) Supervisor - Head of examining committee: Prof. Dr. Şehnaz ŞAHİNKARAKAŞ

....

(University-Inside) Member of examining committee: Dr. Seden TUYAN

....

(University- Outside) Member of examining committee: Dr. Figen YILMAZ

(Çukurova University)

I certify that this thesis conforms to the formal standards of the Institute of Social Sciences.

....

28/09/2018

Assoc. Prof. Dr. Murat KOÇ

Director of the Institute of Social Sciences

Note: The uncited usage of the reports, charts, figures, and photographs in this thesis, whether original or quoted from other sources, is subject to the Law of Works of Art and Thought No: 5846.

DEDICATION

To my father, Selahattin SARI.

ETHICS DECLARATION

Name & Surname : Songül Yeliz SARI

Number : 20118016

Department : English Language Education

Program : English Language Education

Master Thesis (√) Ph.D. Thesis( )

Thesis Title: The Impact of Institutional Tests on Teachers' Teaching
Methodologies

Student's

I hereby declare that;

I prepared this master thesis in accordance with Çağ University Institute of Social
Sciences Thesis Writing Directive,

I prepared this thesis within the framework of academic and ethics rules,

I presented all information, documents, evaluations and findings in accordance with
scientific ethical and moral principles,

I cited all sources to which I made reference in my thesis,

The work of art in this thesis is original,

I hereby acknowledge all possible loss of rights in case of a contrary circumstance (in
case of any circumstance contradicting with my declaration).

28 / 09 / 2018

Songül Yeliz SARI

ACKNOWLEDGMENTS

ABSTRACT

THE IMPACT OF INSTITUTIONAL TESTS ON TEACHERS' TEACHING METHODOLOGIES

Songül Yeliz SARI

Master of Arts, Department of English Language Education
Supervisor: Prof. Dr. Şehnaz ŞAHİNKARAKAŞ
September, 2018, 151 Pages

The aims of this study were twofold. The primary aim was to investigate the washback effects of institutional progress tests on teachers' teaching methodologies. The secondary aim of the study was to explore the role of teachers' pedagogical beliefs and their perceptions of the tests in the mediation of washback. Taking into account the critical role of contextual factors in the generation of washback, a qualitative multiple case study approach was adopted. The study was conducted in the English preparatory school of a university in Turkey, and the participants consisted of 3 teachers working at the given institution. Data were collected in three interrelated stages using the repertory grid technique, classroom observations and in-depth semi-structured interviews. The repertory grid data were analyzed using cluster analysis. The classroom observation and interview data were analyzed using content analysis. The results revealed that the tests exerted washback effects on the teachers' teaching methodologies as well as on teaching content. However, it was found that the tests themselves were not the primary cause of washback. Instead, the results indicated that washback occurred as a result of a complex interplay between the teachers' perceptions of the tests, contextual factors and the teachers' pedagogical beliefs.

Key Words: Washback, Progress tests, Teaching methodology, Pedagogical beliefs, Personal construct theory

ÖZET

KURUMSAL SINAVLARIN ÖĞRETMENLERİN ÖĞRETİM METODOLOJİLERİ ÜZERİNE ETKİSİ

Songül Yeliz SARI

Yüksek Lisans Tezi, İngiliz Dili Eğitimi Anabilim Dalı
Tez Danışmanı: Prof. Dr. Şehnaz ŞAHİNKARAKAŞ
Eylül, 2018, 151 Sayfa

Bu çalışmanın iki amacı vardır. Birincil amacı, kurumsal gelişim sınavlarının öğretmenlerin öğretim metodolojileri üzerine geriye dönük etkisini incelemektir. Çalışmanın ikincil amacı ise öğretmenlerin pedagojik inançlarının ve sınavlarla ilgili algılarının sınavların geriye dönük etkisinin ortaya çıkmasındaki rolünü incelemektir. Bağlamsal etkenlerin sınavların geriye dönük etkisinin ortaya çıkmasındaki kritik rolü göz önünde bulundurularak, çalışma niteliksel çoklu vaka çalışması şeklinde tasarlanmıştır. Çalışma Türkiye'de bir üniversitenin İngilizce hazırlık okulunda yapılmıştır. Çalışmanın katılımcıları söz konusu kurumda çalışan 3 öğretim görevlisinden oluşmaktadır. Veriler, birbiriyle ilintili üç aşamada, repertuar çizelgesi tekniği, sınıfiçi gözlemleri ve yarı-yapılandırılmış derinlemesine görüşme yöntemiyle toplanmıştır. Repertuar çizelgesi verilerinin analizi kümeleme analiz yöntemiyle; sınıfiçi gözlemleri ve görüşmeler aracılığıyla toplanan veriler ise, içerik çözümlemesi tekniğiyle analiz edilmiştir. Sonuçlar, sınavların öğretim içeriğinin yanı sıra, öğretmenlerin öğretim metodolojileri üzerine de geriye dönük etkileri olduğunu göstermiştir. Ancak, bulgular, bu etkilerin ana sebebinin sınavlardan ziyade, öğretmenlerin sınavlarla ilgili algıları, bağlamsal etkenler ve öğretmenlerin pedagojik inançları arasındaki karmaşık bir etkileşim sonucu ortaya çıktığına işaret etmektedir.

Anahtar Kelimeler: Geriye dönük etki, Gelişim sınavları, Öğretim metodolojisi, Pedagojik inançlar, Kişisel yapılar kuramı

PREFACE

Crudely, the teaching and learning process consists of the things that the teacher and learners do for a common goal. Teachers are given the authority to lead the process and entrusted with the duty of remedying the learners' weaknesses; however, bafflingly enough, when it comes to testing and assessing, teachers are often declared inefficient and are urged to keep off testing and assessment by scholars, which contradicts the clichéd statement that "testing and assessment is a natural part of the teaching and learning process." Similarly, in situations where objectivity or maintaining standards is given paramount importance, teachers are often either partially or completely left out of the testing and assessment part of the teaching and learning process. As a result, tests and assessments exert influence not only on test-takers but also on teachers, who are left having to dance in the dark.

Despite all doubts expressed and despite all scholarly efforts to disprove the existence of washback, research to date strongly suggests that where there is a test, there are washback effects. In fact, it stands to reason that when tests carry heavy consequences for teachers, they are very likely to result in washback on teaching. However, research has also shown that not all aspects of teaching are influenced to the same extent and that there may be various factors that play a role in the generation of washback. Since it is a fact of life that testing will continue to be a part of our lives in some shape or form for a long time to come and because there will always be situations where teachers do not have complete control over testing and assessment, it is essential to gain an understanding of these factors in order to minimize the harmful effects of tests, testing and assessment and to maximize their potentially beneficial effects on teaching.

TABLE OF CONTENTS

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

LIST OF ABBREVIATIONS

BA : Bachelor of Arts

CLT : Communicative Language Teaching

ELT : English Language Teaching

ESP : English for Specific Purposes

IELTS : The International English Language Testing System

L & R : Listening and Reading

MA : Master of Arts

TEOG : The Transition Examination from Primary to Secondary Education

TGAT : Task Group on Assessment and Testing

TOEFL : Test of English as a Foreign Language

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF APPENDICES

CHAPTER I

1. INTRODUCTION

This chapter first introduces the background of the study. Then, it presents the statement of the problem, the purpose of the study, the significance of the study, the research questions and the limitations of the study. Finally, relevant key terms and concepts are defined as they are used within the context of this present study.

1.1. Background of the Study

Testing is not a natural but an inevitable part of life in the contemporary world. Many people are faced with various types of tests for a variety of reasons at different stages of their lives. Tests do matter because regardless of their type, their format and what they test, all tests are indeed primarily used as decision- making tools. Thus, "there are always consequences" for test-takers (Cheng & Curtis, 2012, p. 89). It is neither uncommon nor surprising, then, for test-takers to worry about tests and to focus their efforts on test content and test-taking strategies when they believe important decisions will be made about them based on their test scores. Teachers, on the other hand, are not immune to the consequences of tests, either, especially in cases where tests are used as policy-making tools or to standardize teaching (Alderson & Banarjee, 2001; Madaus, 1988; Shohamy, 2007). Research has shown that when a test is externally imposed for such purposes, teachers are very likely to focus their teaching on the requirements of the test in an attempt to deal with the accountability pressures that they face (Burrows, 2004; Cheng, 1995; Kılıçkaya, 2016; Madaus, 1988; Olovsson, 2015; Qi, 2007; Sevimli, 2007; Shohamy, 2007; Spratt, 2005; Şentürk, 2013; Wang, 2010). In the testing and assessment literature, such effects of tests on teaching and learning are usually termed washback.

Although the very existence of washback was questioned by Alderson and Wall in 1993, extensive research has been carried out since then, and it has been clearly shown that washback does exist, especially in situations where the stakes are considered to be high. What has been established just as well is the fact that washback is a complex process which is inevitably shaped by the stakeholders of any given test. There is a general consensus in the literature that among these stakeholders, teachers play a

particularly important role in the mediation of washback due to their critical role in shaping the teaching and learning process.

Research to date suggests that tests cannot induce more washback on teaching than teachers allow them to. A plethora of studies have been conducted to investigate the influence of tests on teaching, and most of these studies have found a direct link between teaching content and test content. However, many studies have yielded mixed results regarding washback on teaching methods and methodologies. The results of these studies suggest that teachers' teaching methods and methodologies are less likely to be influenced by tests. However, it is not clear why teachers are often reluctant to adapt their teaching methodologies in response to tests. It is often suggested that it is crucial to understand why teachers do what they do in order to have a clearer understanding of the mechanism of washback with respect to teaching methodology (Alderson, 2004; Alderson & Hamp-Lyons, 1996; Spratt, 2005; Wall & Alderson, 1993; Wall & Horak, 2011; Wang, 2010; Watanabe, 2004a).

In order to grasp why teachers teach in the ways they do in response to an exam, research to date suggests that it is essential to identify their beliefs about teaching and learning and their perceptions of the test (Alderson, 2004; Alderson & Hamp-Lyons, 1996; Spratt, 2005; Wall & Alderson, 1993; Wall & Horak, 2011; Wang, 2010; Watanabe, 2004a). Surprisingly, however, one would first have to decide what is meant by teachers' beliefs. It is difficult to define the concept of teachers' beliefs based on the current literature because "teachers' beliefs" is still a fuzzy concept, which lacks a clear definition. However, Kelly's theory of personal constructs (1955, 2003) offers a way out of the terminological and conceptual quagmire regarding teachers' beliefs, and within the frame of this theory, teachers' beliefs can be defined as teachers' unique, personal mental representations of ideal teaching and learning practices which have developed as a result of their own experiences. According to Munby (1981), "the significance of teachers' beliefs… to our understanding of teacher decision-making and teacher thinking cannot be overemphasized" because beliefs underpin every educational decision that teachers make (p. 26). Identifying teachers' perceptions of the test in question is also significant not only because a teacher's understanding of test demands plays an important role in shaping the decisions she makes but also because teachers' own perceptions may lead them to consider a test as a high-stakes or a low- stakes test (Hughes, 1993; Madaus, 1988). Madaus (1988) argued that "the power of tests and examinations to affect individuals, institutions, curriculum, or instruction is a perceptual

phenomenon; if students, teachers or administrators believe that the results of an examination are important, it matters very little whether this is really true or false" (p. 35). He explained that irrespective of the actual nature of a test, accountability pressure may cause teachers to regard a given test as high-stakes and thus influence their teaching behavior.

The existing literature clearly suggests that washback is not a teacher-proof formula for changing educational practices. Although most teachers often almost readily adapt teaching content to match the content of a given test, they often vary in their responses with respect to their teaching methodologies. Also, overall, teachers are usually less willing to give up their preferred methods of teaching to meet the requirements of a test. It is generally believed that teacher-related factors such as teachers' personal beliefs about teaching and learning as well as their perceptions of the test lie at the core of such resistance. Thus, it is crucial to analyze such factors in order to understand the nature of washback on teaching methodology.

1.2. Statement of the Problem

Externally-imposed tests are often implicitly or explicitly used as policy-making tools or to standardize educational practices (Alderson & Banarjee, 2001; Madaus, 1988; Shohamy, 2007). At the macro level, this is often done by governments, and tests that serve such purposes include national and international tests developed by professional test designers (Popham, 2003). At the micro level, within individual institutions, such testing practices may include in-house institutional tests as well as national and international tests.

In terms of language testing, in-house institutional tests are not uncommon in Turkey. Many different types of educational institutions such as universities and private language schools create their own assessment systems and tests often with a view to ensuring objectivity and maintaining established standards or in order to reach the institution-set target standards. Contradictory as it sounds, in-house institutional tests function like other externally imposed tests within a given educational institution when they are designed by a designated group of teachers and the teachers that teach the relevant courses are not involved in the test designing process. Such tests are external to the individual teachers who are not involved in the relevant decision-making processes obviously because of the fact that they have no control over the decisions that are made.

Undoubtedly, lack of control over an exam presents teachers with certain challenges in most situations, and teachers' attempts to deal with such challenges may result in the generation of washback (Madaus, 1988).

This present study addresses issues related to the possible effects of institutional testing procedures on teachers' teaching methodologies within the context of the English preparatory school of a university in Turkey. English is the medium of instruction in most of the departments at the university in question, and the English preparatory department offers a one-year-long English program at three different levels which aims to prepare students for their studies at the faculties. The department has its own unique assessment and scoring system. The assessment system broadly consists of a large number of course-specific progress tests including pop quizzes, a number of performance assessments and one general achievement test which is administered at the very end of the year. The progress tests and performance assessments are designed and administered on a course and level-specific basis. However, all the classes within the same level take the same tests at the same time. This highly complex and multilayered system is led by the group coordinators under the supervision of the administration.

The system is heavily laden with exams, and in particular, there are so many progress tests that it seems as if the students are simply and only tested on a regular basis without the opportunity for teachers to help them remedy their weaknesses. The constant testing creates time pressures as most of the tests are administered during teaching hours. Moreover, teachers, apart from those assigned to write the tests, are not informed about test content or the dates of the pop quizzes. As a result, teachers often find themselves stuck in a guessing game and struggling to keep up with the pacing schedule, which is suggestive of the potential of the tests to induce washback.

It is interesting, however, that these progress tests can become a source of concern for teachers and may have the potential to induce washback, given the fact that they are low-stakes tests. As a matter of fact, each individual pop quiz is very low-stakes in nature in terms of the weight given to it. Based on the widely accepted hypothesis put forward by Alderson and Wall (1993) that "tests that do not have important consequences will have no washback," these progress tests, especially each individual test, would not be expected to exert any washback (p. 120). However, on the face of it, the realities of the system suggest otherwise. The truth is, classroom tests, be it externally imposed tests or teacher-made classroom tests, have received very little attention from washback researchers due to this generally held assumption that low-

stakes tests are unlikely to induce washback, and much less is known about whether and to what extent any type of classroom test has the potential to influence how teachers teach.

1.3. Purpose of the Study

The aims of this study are twofold. The primary aim is to investigate the washback effects of the institutional progress tests administered at the English preparatory school of a university in Turkey on teachers' teaching methodologies. The secondary aim is to explore the role of teachers' pedagogical beliefs and their perceptions of the tests in the mediation of washback.

The assessment system used at the institution that forms the setting of this study mostly consists of a large number of progress tests. Indeed, these tests are administered so frequently and are such a great part of the teachers' and students' lives at the institution in question that they often receive much more attention than does the end-of-year-examination until there are no more progress tests left to be administered. Thus, this present study focuses on the influence of the progress tests rather than the end-of-year examination or any other type of assessment tool used at the institution.

1.4. Significance of the Study

A plethora of studies have been conducted to investigate the influence of tests on teaching and learning since Alderson and Wall's call for empirical research (1993). Because of the central role teachers play in the educational process, many of these studies from different parts of the world have particularly focused on the influence of tests on teaching practices. These research efforts have provided convincing evidence that teachers in general do not hesitate much about adapting teaching content to match test demands. However, it has been shown that teachers are much less likely to change their teaching methods and methodologies. It is often suggested that in order to understand teachers' unwillingness to adapt their teaching methodologies in response to an exam, it is essential to examine their own beliefs and perceptions (Alderson, 2004; Alderson & Hamp-Lyons, 1996; Spratt, 2005; Wall & Alderson, 1993; Wall & Horak, 2011; Wang, 2010; Watanabe, 2004a).

Another important component of washback studies includes classroom observation. Based on the understanding that washback is what teachers and students do because of the test, but "would not necessarily otherwise do" (Alderson & Wall, 1993, p. 117), washback on teaching methodology can specifically be defined as the methods and techniques that teachers use to teach because of the test, but would not otherwise use. Because teachers actually use teaching methodology in the classroom, the best way to understand whether a test induces washback on teaching methodology would naturally be to see them in action in the classroom. More importantly, however, research suggests that what teachers say they do and what they actually do may be different (Alderson & Wall, 1993; Farrell, 2015; Lankshear & Knobel, 2004; Stiggins & Conklin, 1992). Thus, it is necessary to see whether what teachers say they do "is reflected in their behavior" in the classroom in order to establish the washback effects of a test (Alderson & Wall, 1993, p. 127). However, despite the increasing recognition of the necessity of classroom observation in washback research, such studies are still outnumbered by those which rely solely on the perceptions and self-reports of teachers.

Much rarer are studies on washback from classroom tests. It is generally assumed that classroom tests are unlikely to induce washback due to the low-stakes involved. However, it is often ignored that "the cumulative effect of a number of classroom tests can lead to a final score that can eventually high stake decision as well" (Shohamy, 1998, p. 344). It is also important to note that the influence of a test is often shaped and determined by the perceptions of those who are affected by the test, and "if students, teachers or administrators believe that the results of an examination are important, it matters very little whether this is really true or false" (Madaus, 1988, p.35). Thus, even if a test is actually low-stakes in nature, if the stakeholders believe it will have important consequences for them, it may influence educational practices.

Overall, worldwide, very little research has been devoted to the investigation of the influence of classroom tests on teaching methodology, and very few of the rare studies on the washback effects of classroom tests on teaching have involved classroom observations or an examination of teachers' beliefs. This gap in the literature worldwide is indeed much larger in the Turkish washback literature, which mainly seems to stem from the fact that most Turkish washback studies have focused on the learner's perspective rather the teacher's. This present washback study focuses particularly on classroom tests, involves an examination of teachers' beliefs and classroom observation and interviews with the teachers as well, and thus aims to fill an important gap in the

literature. It is also assumed that the findings might be beneficial in reviewing and revising the examination system at the institution where the study was conducted. Since these exams constitute an integral part of the examination system used at the institution, such changes would be expected to contribute greatly to improving teaching and learning practices and thereby to help enhance the overall quality of education provided at the institution.

1.5. Statement of Research Questions

The study addresses the following research questions in relation to its purposes:

1) What are the washback effects of the institutional progress tests administered at the English preparatory school of a university in Turkey on teachers' teaching methodologies?

2) What role do the teachers' pedagogical beliefs and their perceptions of the progress tests play in the generation or inhibition of washback on their teaching methodologies?

1.6. Limitations of the Study

This is a qualitative multiple-case study, which was conducted in one institution and only with three participants. Thus, the results cannot be used for generalization purposes. Also, test validity is beyond the scope of this study. Therefore, the study does not involve an analysis of the validity of the tests in question, but relies on the descriptions of the teachers.

1.7. Operational Definitions

Washback: The influence of tests on teaching and learning (Alderson & Wall, 1993).

Test validity: A characteristic of a test, which refers to the capacity of the test to assess what is needed and intended to be assessed (Alderson, Clapham & Wall, 1995).

Construct validity: A unified concept of validity which refers to how well test scores reflect the actual level of skill that the test intends to assess (Messick, 1996).

High-stakes test: A test which has important consequences for the test-taker such as achievement tests, proficiency tests and university entrance examinations (Madaus, 1988).

Low-stakes test: Based on the definition of "high-stakes test" made by Madaus (1988), in this study, a low-stakes test refers to a test which does not have important consequences for test-takers.

Personal construct: A personal assumption or a theory about somebody or something which guides one's behavior toward that person or thing, or similar people or things (Kelly, 1955; Kelly, 2003).

Pedagogical belief: Within the framework of personal construct theory, in this study, a "pedagogical belief" refers to a teacher's personal constructs about ideal teaching practices.

Teaching approach: A teachers' beliefs and assumptions about language learning and teaching (Anthony, 1963).

Teaching method: A general plan for systematic presentation of language that is based on a particular approach (Anthony, 1963).

Teaching/Instructional technique: An exercise, activity or task that a teacher actually uses in the classroom to achieve immediate instructional objectives (Anthony, 1963; Brown, 2001).

Teaching methodology: Actual classroom practices of a teacher; all the techniques that teachers actually use in the classroom to achieve their instructional objectives (Kumaravadivelu; 2006; Thornbury, 2006).

Communicative language teaching: Communicative language teaching is an approach to language teaching that is based on the understanding that grammatical competence by itself is not sufficient to be able to communicate and focuses on communicative competence, instead. It holds that language is learned through meaningful communication and aims to teach real-life language. Learners are viewed as different individuals with different needs and interests and the teacher as a facilitator of learning. Thus, common characteristics of classroom practice of communicative language teaching include: an overall learner-centered approach; the integration of language skills and the teaching of sub-skills; focus on fluency and function as well as accuracy and form; meaningful communicative activities; contextualized presentation of language; use of authentic materials; emphasis on pair and group work; efforts to make learning relevant; inductive as well as deductive teaching of grammar; and teacher

tolerance of errors, especially during fluency activities (Hall, 2011; Richards, 2006; Richards & Rodgers, 2001).

Traditional Approach: Traditional approach to language teaching refers to an approach to language teaching that gives precedence to grammatical competence. Communication is not the focus of instruction, and the main goal is to enable learners to produce grammatically accurate sentences. Immediate error correction is of particular importance lest such errors become habitual and a permanent part of a learner's language. The teacher is viewed as the authority in the classroom, and thus it is a teacher-centered approach. Classroom practices of a traditional approach include teacher-centeredness, focus on grammar and accuracy, deductive teaching of grammar, immediate correction of errors and repetitive practice and drilling (Larsen-Freeman, 2000; Richards, 2006).

Experiential Learning: Experiential learning is a theory of learning that combines theory with hands-on experience in real-life settings and places particular emphasis on the differences between individual learners. In its simplest form, in experiential learning lessons, learners are provided with opportunities for both direct experience and the necessary academic knowledge, and are encouraged to reflect on their experiences. (Kolb & Kolb, 2017).

CHAPTER II

2. LITERATURE REVIEW

2.1. Introduction

This chapter presents the relevant literature. First, it presents the relevant concepts and then a detailed review of the studies conducted around the world and in Turkey.

2.2. Testing and Assessment

In the broadest sense, testing and assessment refers to various procedures used to identify a person's level of knowledge, skill or understanding in a certain area. As in other disciplines and situations which involve some type of testing and assessment, in the field of education and hence in English Language Teaching (ELT), it may easily be observed that the two words, more often than not, stand side by side, creating the impression that they represent a single, unified concept or term. However, in addition to the self-evident fact that these are essentially two different words, the widespread practice of using these two words one after the other in a seemingly random fashion, without specifying a particular reason or explanation causes some confusion as to what is actually meant by testing and assessment. Therefore, before all, there exist three crucial questions to be answered: 1) What precisely is testing? 2) What precisely is assessment? 3) Is there a difference at all?

At first glance, it is definitely not too difficult to be tempted into assuming that the two words are used interchangeably:

"To begin with, the very nature of testing has changed quite radically over the years..., conceived not so much as to catch people out on what they do not know, but as a more neutral assessment of what they do" (McNamara, 2000, p. 4).

Similarly, one could cheerfully cease to focus on testing as a derivative of test when she/he finds out that tests may be used "as a means of assessing the student's performance" (Heaton, 1990, p. 5). It would not be really fair to blame the practicing teacher with no expert knowledge for giving up pondering over the difference considering that assessment literacy is a worldwide problem and "the increasingly central role of testing and assessment" requires a higher degree of "assessment literacy"

(Taylor, 2009, p. 25). Much as Brown (1996) might cause serious unease and questioning due to his adherence to the word measure most of the time when he refers to tests, the reader might, misguidedly, feel relieved by the fact that different types of subtests assess different skills in different ways. After all, teachers use tests in the classroom to "assess their students" (Shohamy, 1998, p. 344). Shepard's (1994) statement that test and assessment actually "mean the same thing" and the difference is "of symbolic importance" helps reinforce the perception that *assessment* and *test* are synonyms (p. 206).

The existing literature is also open to the interpretation that testing refers to traditional testing consisting of mainly paper-and-pencil tests or large-scale tests whereas assessment to classroom assessment, which, to make matters more confusing, is often classified as formative assessment. One example is Fulcher and Davidson (2007), who, after pointing out the difference between large-scale tests and the classroom environment, go on to explain classroom assessment. They underscore that classroom assessment is a long, ongoing process and therefore formative. Shepard (2000) makes an even clearer distinction as she criticizes the exceptional power attached to educational tests. She laments that assessment was ignored for a long time as an important part of the learning process and that neither classroom assessment nor "its traditional counterpart, tests and measurement" received enough attention from researchers (p. 1). She also draws a contrast between external and classroom assessment, as the former serving summative purposes and the latter formative. Likewise, Hughes (2003) mentions the different purposes of large-scale, standardized tests and teachers' informal assessments, which are formative in nature.

There is no doubt that testing and assessment are interrelated. Nor is it possible to deny the formative functions of classroom assessment or the summative functions of large-scale, standardized tests as well as paper-and-pencil tests. However, a closer and patient examination of the literature reveals that there is actually more to it than meets the eye. There are differences and there are points of convergence between testing and assessment, and their summative and formative functions, which may not be obvious at first sight. Therefore, in order to be able to have a clearer understanding of testing and assessment, it may be more helpful to consider certain subtle details before reaching quick conclusions.

Perhaps the most important fact that needs to be taken into account is that tests are within the scope of assessment. Put more simply, tests are a type of assessment.

According to the TGAT report, which Black (1998a) cites in objection to Shepard's (1994) notion that tests and assessment are one and the same, assessment is an umbrella term "embracing all methods customarily used to appraise performance of an individual pupil or group" (Department of Education and Science, 1988, as cited in Black, 1998a, p. 5). It is also stated that "an assessment instrument may be any method or procedure, formal or informal for producing information about pupils: for example, a written test paper, an interview schedule, a measurement task using equipment, a class quiz" (Department of Education and Science, 1988, as cited in Black, 1998a, p. 5). Similarly, Berry (2008) defines assessment as the act of collecting information about student learning through different means, including tests. In this sense, the relationship between assessment and tests may be likened to that between the whole and the parts. Tests are among the many parts, in this case the many different methods, which make up the whole, that is, assessment. Therefore, it seems that test and assessment are neither synonyms nor a binary opposition.

This understanding of assessment as an all-encompassing term to describe different possible methods that may be used to collect information about students and student learning certainly helps clear up the mystery regarding how tests can assess students. It becomes obvious that tests are literally a means of assessment. However, a test is distinctly different from other forms of assessment in that it is "conducted within formal and specified procedures, designed to ensure comparability of results between different test administerers and between different test occasions" (Department of Education and Science, 1988, as cited in Black, 1998a, p. 5). Another definition with the advantage of brevity belongs to Berry (2008): "A test often results in numbers." (p. 6). She explains that a test is a systematic procedure which aims to describe a person's level of skill or knowledge in a given area by assigning a number to it based on a certain scoring system and a set of rules, which is called measurement. Hancock (1994) also emphasizes that tests are formal assessment procedures which sample and measure student learning. Simply put, tests are assessment methods used for scoring or grading student learning based on certain criteria.

With some clear definitions of test and assessment, the testing/ assessment conundrum slowly begins to unravel. Obviously, testing and assessment are closely connected, but at the same time, they mean different things. Based on the fact that tests translate student learning into numbers, testing can simply be considered as the specific practice of assigning a number to one's knowledge of or ability in a certain area.

However, assessment is a general term, which refers to methods for collecting information about student learning. It comprises tests and other methods which do not necessarily involve any scoring or grading. Thus, it is not easy to make a clear distinction between the summative and formative purposes of tests and assessment. It should be borne in mind that there are different types of tests and different forms of assessment with different purposes. It is essential to know what these different kinds of tests and assessments involve if we wish to fathom out their effects on teaching and learning.

2.3. Types of Tests

It is possible to derive a variety of categories from the literature regarding the different types of tests. Such categories may range from tests grouped by test format, test method and approaches to testing to tests grouped by test purpose and score interpretation. However, it should be noted that technical details of tests are beyond the scope of this study. Therefore, in accord with its purpose, the present study focuses mainly on the classification of tests by purpose, which broadly refers to the reason why a particular test is administered (Heaton, 1990; McNamara, 2000). According to this approach, tests often fall into four major groups: placement tests, progress tests achievement tests and proficiency tests.

2.3.1. Placement Tests

Placements tests are often used by various language institutions prior to instruction in order to determine students' levels of proficiency and to place them into appropriate groups (Brown, 1996, Coombe, Folse & Hubley, 2007; Hughes, 2003). They are mainly administered to decide at which level a student should start her/his language training to accomplish specific program goals. Therefore, they are program-specific tests used for program-level decisions and are often designed by individual institutions (Brown, 1996; Hughes, 2003). It is possible to use a general proficiency test for placement purposes if a given program includes the wide range of levels that proficiency tests normally cover. However, if a program is limited in terms of the levels it comprises, it is necessary to base the placement test on the specific purposes of the program and the curriculum it follows (Brown, 1996; Hughes, 2003).

2.3.2. Progress Tests

Progress tests are tests which are used to monitor student progress. They measure student learning in relation to individual instructional goals, which are parts of the overall intended learning outcomes of a particular course. Accordingly, these tests are given at relatively short intervals at the end of each instructional unit with specific objectives (Brown, 2003; Coombe et al., 2007; Hughes, 2003). As such objectives are both defined and implemented by teachers, progress tests are usually designed by teachers themselves. Clearly, the major function of a progress test is to provide information about how successfully the process is progressing towards the ultimate goal based on the specific features that characterize this process (Heaton, 1990; Hughes, 2003). Thus, in addition to revealing how much learning has been achieved, progress tests, as a mirror of the process, play an important role in shaping teaching and learning on a classroom level.

For teachers, progress tests are a means of assessing both their students' performance and their own. The results of a progress test help teachers form a clear opinion of the degree to which their students are achieving specific learning goals, based on a tangible piece of evidence. Such evidence can be used as a method for grading achievement in the long run. However, more importantly, it enables teachers to focus on the areas their students have difficulty in and thus to decide which aspects of their teaching may need to be reviewed. Teachers can thereby make the necessary adjustments to their teaching in order to remedy weaknesses and enhance learning (Heaton, 1990; Hughes, 2003).

The most important influence of progress tests on students is related to motivation. Being tested at regular intervals provides students with a clear goal to work towards. Achieving high scores increases confidence and helps the students maintain their focus and interest whereas low scores may promote more effort. In either case, what is important is that students are informed about their progress on a regular basis rather than being left in the dark until the end of the semester or year. Considering that progress tests are based on what has been covered in class, such information would be expected to lead to improving scores provided that it is supported with timely and appropriate feedback (Black & William, 1998b; Heaton, 1990).

2.3.3. Achievement Tests

It is possible to consider achievement tests as the one-shot version of progress tests. Like progress tests, an achievement test measures student learning based on course objectives and the curriculum (Brown, 1996; Coombe et al., 2007; Heaton, 1990; Hughes, 2003). The purpose is to determine the level of learning achieved as a result of instruction, and the results may be used as a means of identifying certain weaknesses in teaching and/or learning (Brown, 1996). However, the major difference between a progress test and an achievement test is that the former is used to make decisions about the process over the course of the process whereas the latter is used to make formal decisions regarding whether or not the student has attained the knowledge or skills necessary to move to a more advanced level of study. Therefore, achievement tests have a larger focus and are normally administered at the end of instruction (Brown, 1996; Hughes, 2003).

Based on this information, it is possible to classify achievement tests into two broad categories, which consist of end-of-course tests produced by course instructors and large-scale, standardized tests designed by professional examining bodies and administered by the state (Heaton, 1990; Hughes, 2003). Both teacher-produced and state-administered achievement tests are used for making promotion decisions. Test scores are assumed to denote success or failure, that is, the achievement of course objectives or lack thereof. Students may pass or fail, or graduation may be dependent upon the scores obtained from such tests (Brown, 1996; Heaton, 1990). In addition to promotion decisions, state-administered tests are often used for accountability purposes. Schools may be ranked according to success rates and teachers are held accountable for high failure rates, which inevitably puts pressure on teachers and school administrations (Heaton, 1990; Popham, 2003). These important consequences may cause achievement tests to exert a considerable influence on teaching and consequently on learning (Heaton, 1990).

2.3.4. Proficiency Tests

According to the Oxford Advanced Learner's Dictionary, proficiency means the ability "to do something well because of training and practice" (Oxford University Press, 2018). In terms of language, that broadly translates as the ability to use a

language correctly and efficiently. More specifically, it refers to having an adequate command of a language, which enables a non-native speaker of the language to communicate in that language with ease in a particular setting. An individual who has the required knowledge and skills is considered proficient (Hughes, 2003). Whether or not a person is proficient in a language is determined by using proficiency tests.

As opposed to the aforementioned types of tests, proficiency tests are not specific to any particular teaching situation, course or program. They are more general and comprehensive tests which measure individuals' overall knowledge of and ability in a language independently of any curriculum, course or program content and are often developed by professional, external examining boards (Alderson, Clapham & Wall, 1995; Brown, 1996; Coombe et al., 2007; Heaton, 1990; Hughes, 2003). The purpose of a proficiency test is not to assess how much learning has been achieved with respect to what has been taught in a particular course, but rather to determine how strong the candidate's command of a particular language is irrespective of her/his language learning background (Alderson et al., 1995; Heaton, 1990; Hughes, 2003).

2.4. Alternative Assessment

The major promise of alternative assessment is to support learning. Based on the understanding that learning is a process of constructing knowledge, in this approach, assessment is mainly used as a means of facilitating this process rather than a measurement tool that determines how much knowledge has been acquired. Therefore, the learning process itself is of great importance, and before a final decision is made, students are assessed throughout the process based on the objectives that guide this process. The assessment tools used for assessing the process include interviews with students, documented observations, student learning logs and journals, graphic organizers, checklists and student self-evaluation (Scott, 2000; Stiggins, 1992). The common distinguishing feature of these tools is that they provide feedback which focuses on individual students and guides them through the process rather than only reveal the common weaknesses and strengths of a group of learners. That kind of information is valuable in that it helps both the teacher and the learner develop customized strategies to enhance learning, and it thus increases the likelihood of student success. (Hamayan, 1995; Scott, 2000; Stiggins, 1992).

2.5. Classroom Assessment

Classroom assessment, as the name suggests, refers to assessment carried out in the classroom. Although at first sight it seems like a superfluous term since most of formal education and hence assessment typically takes place in the classroom, it is mainly this very fact that makes it a meaningful and important concept. Primarily, assessment is a thing of the classroom and thus a natural part of the teaching-learning process. Throughout this process, from beginning to end, teachers constantly assess their students in order to check their understanding and to gauge their abilities. Following that, based on the information they have obtained, they make a series of decisions. Such decisions include those whose purpose is mainly to improve learning and those which are mainly used as tangible evidence of student learning. Thus, classroom assessment entails scoring and grading student learning as well as assessing the process to improve learning (Berry, 2008; Leahy, Lyon, Thompson & William, 2005).

2.6. Formative or Summative?

In the field of testing and assessment, the adjectives formative and summative are used to describe assessment in terms of the purpose for which teachers assess their students. In other words, they refer to the function the assessment serves. Assessment is considered formative if it is integrated into teaching with a view to supporting and improving learning. Such assessment normally runs in parallel to learning and involves monitoring the progress of students in order to help them successfully complete the process. The word, summative, on the other hand, refers to the grading function of assessment. The main goal is to report to third parties how much learning the student has achieved at the end of the process, which requires the teacher to make a final judgment and to fit it into a recognizable symbol (Berry, 2008; Chappuis, 2009; Stiggins, 2000). Although the summative use of assessment evokes unfavorable feelings in many, the fact remains that the nature of formal education makes it mandatory. Therefore, the question is not whether students should be assessed summatively, but rather how the two facets of assessment can be linked to each other so that assessment can promote learning. Considering the benefits of formative assessment, the simple answer seems to lie in following a similar path when it is time to make summative

decisions, that is, diffusing summative assessment into the process instead of delaying it to the very end and using summative assessment tools for informative purposes. When used in this way, summative assessment not only provides more reliable information about student achievement at the end, but also takes on a formative role and can thus contribute to the successful completion of the process (Berry, 2008).

2.7. Washback

Exams are formal assessment tools, whose primary purpose is to reveal information about student learning. A series of important formal and informal decisions are made based upon this information, which may not only influence the course of the process, but also change the future lives of students. Exams have thus long been the focus of many researchers and scholars. Extensive research has been conducted in order to improve examination methods and to render exams reliable sources of information for decision-makers. The increasing use of high-stakes and large-scale exams has added a new dimension to this quest by bringing into focus the impact of exams prior to administration, and washback has thus come to be considered as an important topic of research in the field of testing and assessment. The existing literature includes different approaches and conceptualizations as well as a number of studies based on the relevant theory.

2.7.1. The Definition of Washback

Washback (or backwash) can broadly be defined as the influence of exams on teaching and learning practices (Alderson & Wall, 1993; Heaton, 1990; Hughes, 2003; McNamara, 2000; Messick, 1996). Alderson and Wall (1993) were among the pioneering researchers that have attempted to clarify the concept, and in their skeptical review of the literature, they broadly separate washback from any other possible effects of exams and narrow the definition down to the things teachers and learners do because of the test, but "would not necessarily otherwise do" (p. 117). Drawing on Alderson and Wall's definition, Messick (1996) also defines washback as "the extent to which the test influences language teachers and learners to do things they would not otherwise necessarily do" (p. 1). Some other researchers make a distinction between the classroom and society- level effects of exams and use the term test impact to refer to the broader

and large-scale effects of tests on society (Bachman & Palmer, 1996; Hamp-Lyons, 1997; McNamara, 2000; Wall, 1997). According to Hamp-Lyons (1997), Messick's notion of consequential validity, which embraces washback as one of the "potential consequences of test use" offers the middle ground between washback and impact that is necessary to tackle the ethical problems concerning test design and test use (Messick, 1996, p. 10). In this present study, washback and impact are used interchangeably to refer to the influence of tests on teaching and learning practices.

2.7.2. The Mechanism of Washback

The concept of washback is based on the assumption that exams have the power to shape learning and teaching practices (Cheng & Curtis, 2004). From this perspective, the primary function of an exam should be to yield positive washback effects and eliminate any negative washback effects. A number of researchers have proposed different models of washback, delineating how positive washback can be achieved. Most of these suggestions mainly focus on the importance of improving test validity, test content and test method and linking instructional practices closely to the test. However, recent research has shown that the mechanism of washback is more complex than what shows on the surface and that there may not be a straightforward path to achieving positive washback.

One of the earliest proposals to use exams as a driving force for improved instruction came from Popham in 1987. He argued that the simplest way of repelling the negative effects of high-stakes tests was to adopt a well-designed system of measurement-driven instruction. He formulated a cost-friendly formula, the key elements of which can be briefly summarized as improving exam content, informing teachers about test content and providing teachers with instructional support where necessary. He robustly asserted that transforming exams was the shortest cut to achieving educational improvement. What lies at the core of this easy-on-the-ear solution seems to be a fatalistic attitude to the reality of high-stakes testing. The idea is that if high-stakes tests are inevitable, we should refine them and enjoy the educational benefits. In other words, it is implied that the most viable option is for the testers to improve the system and re-join it since they cannot beat it and it is "almost certainly" guaranteed that teachers, all of them, or at the least the ones in the USA, will do as they are prescribed (p. 680).

Fredericksen and Collins (1989) were essentially of the same mind as Popham. They, too, advocated the use of exams for the purpose of improving educational practices. They believed that a well-designed test had the power to raise teaching standards. Focusing on the educational system as a whole, they developed a broader notion of construct validity, which they called systemic validity. According to this holistic understanding of validity, a systemically valid test is one that improves instruction in a particular educational system, whereas a systemically invalid test is a test that has a negative impact on teaching and learning. Although the writers did not use the term washback, obviously, this approach to test validity is directly related to washback effects. It is suggested that positive washback effects are an indicator of test validity and that high-quality tests would be expected to produce positive washback effects, hence the importance of improving the systemic validity of tests. The solution Frederiksen and Collins (1989) put forward in order to produce systemically valid tests was to evaluate student performance on extended tasks rather than on single-item level. They underscored that it was crucial to adopt such an assessment method in order to measure higher-order cognitive skills and argued that emphasis on such skills on the test would lead to a corresponding emphasis while teaching. Thus, a well-constructed test designed based on this principle would ultimately be expected to bring about improvement in instructional practices and hence in learning in the educational system in which it is used.

In their well-known, comprehensive review of the literature, Alderson and Wall (1993) pointed out the importance of considering what may be called the teacher and the student factors in understanding how washback operates. In fact, they put the teacher and the learner at the heart of the mechanism of washback and defined washback effects as the things teachers and learners do because of the test, but "*would not necessarily otherwise do*" (p. 117). Based on this definition, they argued that any test, whether good or bad, could lead teachers and students to act in certain ways. They explained that poor tests could indeed induce positive effects on both teachers and learners by motivating them and encouraging hard work. Under the same rationale, high-quality tests may cause negative effects by increasing anxiety for both parties. Therefore, they asserted that washback could not be "directly related to a test's validity" (p. 116), and they thus strictly rejected Fredericksen and Collins's notion of systemic validity (1989). In fact, they were highly skeptical about whether washback existed at all. They cautioned against assuming that a test would definitely cause washback and

noted that it was necessary to "take account of different factors" and check for evidence before asserting the existence of washback (p. 120). Based on their rigorous analysis of the literature, they put forward 15 hypotheses of washback in order for researchers to consider before they set out to investigate the washback effects of a test (Alderson & Wall, 1993; pp. 120-121):

(1) A test will influence teaching.

(2) A test will influence learning.

(3) A test will influence **what** teachers teach; and

(4) A test will influence **how** teachers teach

(5) A test will influence **what** learners learn; and

(7) A test will influence the **rate** and **sequence** of teaching; and

(8) A test will influence the **rate** and **sequence** of learning.

(9) A test will influence the **degree** and **depth** of teaching; and

(10) A test will influence the **degree** and **depth** of learning.

(11) A test will influence attitudes to the content, method, etc. of teaching and learning.

(12) Tests that have important consequences will have washback; and conversely

(13) Tests that do not have important consequences will have no washback.

(14) Tests will have washback on all learners and teachers.

(15) Tests will have washback effects for some learners and some teachers, but not for others.

Similarly, Messick (1996) argued that washback effects of a test could not be attributed to the validity of the test. He developed a unified concept of construct validity (1989), which considers washback as one aspect of consequential validity (as cited in Messick, 1996). According to Messick (1996), "washback is only one form of testing consequence…. and testing consequences are only one aspect of construct validity" (p. 2). Therefore, "neither testing consequences in general nor washback can stand alone as a standard of validity" (p. 2). He pointed out that there might exist various other factors leading to undesirable effects on learning, independent of the validity of a test. He explained that although a valid test would be expected to produce positive washback effects and an invalid test could cause negative washback effects, such effects may indeed be the consequences of educational practices themselves rather than of the test.

Hughes's trichotomy of washback (1993, as cited in Bailey, 1996) which consists of participants, process and the product, puts a similar emphasis on how the targets and recipients of washback can indeed shape and mediate washback effects. According to this model, participants broadly include anyone involved in the educational system who is at the recipient end of a given test. The process refers to the actions which the participants take in response to a given test and the product to the result of these actions. He explains that a test may generate washback effects by first influencing the participants' understanding of what is expected from them, which may in turn, lead them to take certain actions in order to meet those expectations. Finally, their actions will influence the product, that is, what is learned and how well it is learned. This model clearly shows that individuals' interpretations of the test and their corresponding actions play an important role in creating washback. This understanding can also be seen in Hughes's approach to generating positive washback (2003). According to Hughes (2003), in order to generate positive washback, it is necessary to (pp. 53-56):

1. Test the abilities whose development you want to encourage.

2. Sample widely and unpredictably.

3. Use direct testing.

4. Make testing criterion-referenced.

5. Base achievement tests on objectives.

6. Ensure the test is known and understood by students and teachers.

7. Where necessary provide assistance to teachers.

Hughes holds that students and teachers play a key role in the mediation of washback, so they inevitably play a pivotal part in the generation of positive washback. He points out that a test cannot exert the intended and expected positive washback effects unless the participants fully grasp what they are expected to do. Thus, he lists clarifying students and teachers about the purpose and nature of the test and lending support to teachers where necessary among the important steps that should be taken to achieve positive washback.

Based on Alderson and Wall's (1993) and Hughes's theories, Bailey (1996) developed a basic model of washback which makes a distinction between washback to the learner and washback to the program. Washback to the learner refers to any action taken or strategy used by the learner with the purpose of being successful on an exam whereas washback to the program refers to the effects of the test on how the language is

taught and how "test-driven information" is used by "teachers, administrators, counselors, curriculum developers, and so on" (p. 264). She suggested that whether a test resulted in positive or negative washback effects depended on whether it fostered or inhibited learning. In her review of the literature, she drew attention to the growing perception of traditional testing as the major cause of negative washback as a result of the changing attitudes to teaching, that is, the shift from traditional teaching methods to communicative language teaching. Pointing out the incongruity between this new approach to teaching and traditional testing and based on the idea that a communicative test should aim for washback, she suggested using communicative tests as a means to achieve positive washback. Drawing on the literature, Bailey concluded that a communicative test which could promote positive washback would: 1) be consistent with educational goals 2) assess real-life skills through lifelike tasks or relevant texts 3) lead the learner to take responsibility for her learning and allow for self-assessment and 4) provide detailed feedback on test performance. Bailey's overall model of washback suggests that such a test may result in the improvement or enhancement of learning and/or teaching activities by influencing the learner and the program.

More recently, Watanabe (2004a) developed a different approach in order to illustrate the complexities of washback. Based on the relevant literature, he conceptualized washback in terms of: 1) its different dimensions such as the specificity, intensity (Cheng, 1995), length, intentionality, and value of washback 2) aspects of learning and teaching that may be influenced by the examination and 3) factors mediating the process of washback being generated. This approach is based on the awareness that washback is a multifaceted phenomenon and not an all-or-nothing concept. It is suggested that washback may exist on different levels, may affect different aspects of learning and teaching, and may result from many different factors apart from the test itself. Given this complexity, it is considered crucial that researchers "take account of the whole context wherein the test is used" (Watanabe, 2004a, p. 22) and incorporate qualitative research methods into their studies in order to be able to understand how and to what extent a given context and the test interact to generate washback (Cheng, 2004).

2.7.3. Washback Studies from Around the World

Alderson and Wall's (1993) critical discussion of the concept of washback laid

the foundations of modern systematic washback research. It brought into sharp focus the different factors that might cause a test to exert influence on educational practices and hence the different aspects of washback. The 15 hypotheses provided a practical diagnostic list of the possible signs and symptoms of washback, which quickly became a popular checklist used to check for evidence of washback in washback studies. Owing to the detailed and inclusive nature of this list, the trend in washback research slowly shifted from a view of washback as a fixed, single-faceted phenomenon toward focusing on the different layers and dimensions of the concept. Now, researchers do not investigate only washback effects of an exam. They investigate or identify washback on teaching, washback on learning, washback on attitudes, and the type and extent of washback among many other things.

Changes in teaching and learning being the core criteria on the list, Alderson and Wall (1993) put a special emphasis on the necessity of finding out what really goes on in the classroom. They strongly argued that this could not be achieved without classroom observations. The writers' criticism of studies relying merely on the accounts of teachers and students in addition to their well-justified call for more studies to meet the crying need for empirical data soon generated an interest in classroom observation. As a result, in terms of washback on teaching and learning behavior, conducting classroom observations became the norm in washback research, and the last two decades have been marked by an increase in the number of studies which have investigated washback on classroom practices through classroom observation in combination with other methods.

Alderson and Wall (1993) indeed pioneered in taking their own advice. Their Sri Lankan study was the first study to examine the influence of testing on classroom practices through classroom observation in combination with questionnaires and interviews with the teachers (Wall & Alderson, 1993). In their study, they investigated the effects of the O-Level English examination, introduced by the Sri Lankan government to ensure the implementation of the changes made to the teaching program via the production and introduction of a new series of textbooks. Since the textbooks were used as the primary agent of the intended improvements in the teaching of English, the exam was designed based on the advice provided by the textbook writers and as Wall and Alderson ascertained (1993), reflected the objectives of the books. Considering this direct relationship of the textbook to the test, in order to be able to have a full and clear picture of the influence of the test, the researchers conducted

baseline studies six months before the administration of the exam for the first time, the results of which revealed that teachers had not yet switched to a communicative methodology although they claimed that they had. The main study was carried out several years after the exam came into use and focused on two groups of classes: 1) where the textbook was used and 2) where other materials were utilized. The results showed that in the classes where the textbook was used, the teachers strictly followed the book in terms of its content; however, their methodology did not match the philosophy of the book although at times they incorporated exam-related tasks into their teaching. In the classes where the textbooks were not used, the teachers used test-related materials, but as was the case with the other group, their overall teaching methodologies were not in line with what the textbook aimed to encourage teachers to do. Given that the test was based on the textbooks, the overall results indicated that the exam exerted a strong washback effect on teaching content whereas it did not have any observable washback on how the teachers taught. It was also found that the test influenced to some extent how the teachers assessed their students, but not how they graded them. Based on these results, Wall and Alderson (1993) concluded that the teachers understood neither the purpose of the exam nor the philosophy of the textbook, and the researchers ascribed it mainly to a lack of training, which indeed seems to be a recurrent theme in their report.

One of the most well-known studies which have followed Wall and Alderson's model (1993) was conducted by Cheng (1995, 1999, 2004). The aim of her well-publicized, dissertational study was to find out whether the revised Hong Kong Certificate of Education Examination in English (HKCEE) had any washback effects on educational practices in Hong Kong secondary schools. She administered questionnaires, and conducted classroom observations and interviews with government officials, textbook publishers as well as with teachers. For comparative purposes, she conducted the teacher survey before and after the examination came into effect and classroom observations of the same teachers teaching two different groups of students: students who were going to take the old HCKEE in 1995 and students who were going to take the revised HCKEE in 1996. The preliminary research findings (Cheng, 1995) indicated that teachers were skeptical and anxious about the new exam whereas they were willing to change their methodology to match the new exam. In terms of their actual classroom practices, these mixed feelings seemed to manifest themselves in the tasks that were taught, which were congruent with the types of tasks students were

going to be asked to perform on the exam. The teachers employed different activities and taught different tasks in the different classes they were teaching depending on which exam (new or old HCKEE) their students were going to take. The results of the full study showed that, in terms of their perceptions of the exam, teachers had developed a more positive attitude towards the exam; however, their interest in changing their teaching methods and methodologies to suit the philosophy of the exam had decreased (Cheng, 2004). In her report, Cheng (2004) explains this difference as stemming from the fact that "what the teachers think they would like to embrace in terms of the new exam and what they can actually do in teaching might not necessarily match" (p. 158). However, further classroom observation supported previous findings (Cheng, 1999). There was a significant change in the way the teachers organized their lessons when teaching students who were going to take the revised exam. Although their teaching mainly followed the same patterns, they allocated more time to pair and group work and focused on exam-related tasks such as role-play and group discussions. Based on Cheng's (1999) observation that there was no change in the teachers' "fundamental beliefs and attitudes about teaching and learning, the roles of teachers and students, how teaching and learning should be carried out" (p. 268), that might be interpreted as a "superficial" or even as a fake washback effect on the teachers' teaching methodologies (Cheng, 2004; p. 163). From a different perspective, however, considering the nature of belief systems, it could also be considered as the very initial symptoms of washback, the development and permanence of which depends on both its future results and the efforts made to support it through practical assistance provided to the teachers. While the findings of this study with regard to teaching methodology is open to different interpretations, there is no doubt that with respect to teaching content, the results point to a strong washback effect as Cheng (1995, 1999, 2004) reported that teaching was heavily based on textbooks and commercially written exam materials specifically designed for the exam(s).

Watanabe (2004b) was also one of the earliest researchers to employ classroom observation to examine the washback effects of a national test. In 1994, he set out to investigate the influence of the English component of the Japanese university entrance examination on how teachers delivered instruction. The study involved 5 teacher participants, who were observed in their regular and exam preparation classes for a total of 964 minutes. The participants were also interviewed before and after the observations took place in order to collect personal information and to discuss the observation

results. Considering the overall nature of the examination, Watanabe predicted that the teachers' teaching would be dominated by a focus on structure and the use of grammar-translation method whereas there would be little opportunity for the practice of aural/oral skills in their exam preparation classes. However, the results revealed unforeseen complexities of teacher behavior. Although the teaching behaviors of some teachers seemed to justify the researcher's assumptions, the fact that they employed those methods more frequently in their regular classes makes it difficult to interpret their behavior as washback. Similarly, as opposed to the predictions, some teachers continued to have their students practice listening although it was not a predominant feature of the exam. However, it is as difficult to dismiss the existence of washback out of hand considering that such practice consisted mainly of mechanical tasks similar to those students were required to perform on the exam. In his report, Watanabe (2004b) lists the teachers' lack of knowledge of alternative methods and their varying perceptions of the exam and testing in general as important factors which may explain these complex results. In addition to these, it is possible to deduce from the interviews with the teachers that the teachers' different beliefs and conceptions regarding teaching and learning might be another factor which may have shaped their perceptions of and thus their reaction to the exam. It is also important to note that although it was not the focus of the study, it can easily be understood from the report that the teachers depended heavily on past exam papers in their exam preparation classes. Thus, it may be concluded that the exam had a direct bearing on what the teachers taught; however, it did not penetrate deep into their teaching methods.

Glover (2006) conducted an unusual comparative study to find out whether the end-of-high school examinations administered in a country in Central Europe exerted any washback effects on how teachers delivered instruction. In his study, he worked with two teachers and compared the teaching methods they used in their exam-preparation classes and the ones they used in their regular classes. Two special features of his study were the exclusive focus on teacher talk as the primary data collection method and the use of discourse analysis for data analysis. In order to complement the primary data, he also conducted a survey, interviews and classroom observations. The results revealed obvious washback effects on teaching content as the curriculum was strictly based on what was covered in the exam and past exam papers. It was also found that the teachers had negative attitudes toward the test due to the reason that they felt as if they had to teach differently from how they would normally teach, apparently,

primarily because of time constraints. The teachers' claim seems to be supported by the results of the discourse analysis, which demonstrated that the teachers had used more direct methods in their exam preparation classes such as noticeably fewer cued elicitation words and correction response words. Simply put, the teachers had spent less time on guiding students and trying to engage them in their own learning, which may be interpreted as an indication of negative washback. However, on the other hand, the teachers were also found to have provided far more explanations and feedback about certain language points in response to student utterances, which may be considered as positive washback owing to the potential of such teacher intervention to help improve learning. In addition to these, an interesting finding was that neither teacher gave up pair/group work although such activities were not part of the exam in question. This could be linked to an underlying belief that such activities may facilitate learning. Similarly, the presence of points of divergence in terms of the degree of change the teachers displayed and the aspects of their teaching that changed may be related to their differing beliefs about teaching. However, it is not possible to carry this argument beyond speculation since there is no clear information provided with regard to the individual beliefs of the teachers. Nonetheless, the overall results of this study seem to support the argument that a test may have positive and/or negative effects regardless of how well-designed it is as a result of the perceptions and behaviors of those who are affected by the test (Alderson & Wall, 1993; Alderson & Hamp-Lyons, 1996; Hughes, 1993; Watanabe, 2004a).

Shih (2009) also conducted a comparative study in order to explore the effects of a national test on teaching and learning. She compared two universities in Taiwan, one of which (University A) did not impose the given test as a degree requirement, whereas the other (University B) did. She observed three teachers in class at University B and two teachers at University A. She also conducted interviews with teachers, students and administrators in order to support the observational data. The results of the classroom observations showed that only one teacher at University B employed test-oriented instructional methods. However, the interviews with the students about some other teachers who were not observed by the researcher actually suggest that there might indeed be a significant difference between the two universities in terms of the washback effects the test exerted on them. Based on these findings, the researcher reported that there was not a striking difference between the two universities regarding the washback effects of the examination, concluding that the test had a limited impact on teaching

practices, which she attributed to teacher-related factors rather than the test itself. The lack of attention given to the accounts of the students in the interpretation of the results could well be due to the rationale that there is no logical reason to leave ourselves to the mercies of student accounts when we are strictly cautioned against relying completely on teacher accounts without having been a witness to their classroom practices ourselves. Regardless, the results of this study are consistent with those of Glover's in that they point to a role of the individual differences of teachers in the generation of washback.

The findings of Qi's study (2007) leaves relatively little room for speculation as regards the potential of teacher-related factors to influence the hoped-for influence of tests on teaching. Her comprehensive research aimed to examine the washback effects of the writing task in the National Matriculation English Test in China (NMET) on pedagogical practices in relation to the intentions of the test designers. She collected data through interviews, questionnaires, classroom observations and examination of test-related documents as well as of teaching materials. The interviews with the test constructors showed that the exam was intended to bring about a positive change in teaching and consequently in learning. It was expected that the strong communicative focus of the writing task in this adapted version of the test would result in a parallel change in teaching. Apparently, the test constructors, with the best of intentions, like many others, had designed the test in the hope that it would prompt teachers to abandon traditional methods of teaching and switch to a communicative approach instead. The examination of the test papers corroborated the test constructors' self-reported intentions. However, the data gained from the teacher and student interviews told a noticeably different story. Despite the fact that the teachers' and students' perception of accuracy, organization and content as being among the most important aspects of writing coincided with that of the test designers, there appeared to be a large gap, not to say a gaping chasm, between the views of the two parties on the communicative aspects of writing. Ironically, the teachers did not even mention anything remotely related to the major communicative features of the new exam, which the test constructors were so eager to spark an interest in. It seemed that the teachers were just as focused on the mechanics as before, if not even more so, at the expense of ignoring the communicative purposes of writing. The results of the student and the teacher questionnaires affirmed the findings of the interviews. Obviously, the classroom observations provided the researcher with a real-time experience of what the interviews and questionnaires had

already found. It was revealed that the teachers were concerned about accuracy the most and paid very little attention to the communicative aspects emphasized by the test constructors. Actually, only one of the four teachers observed seemed to show some interest in teaching the communicative aspects accentuated in the exam. However, sadly, the teacher's well-intentioned yet misguided efforts to teach how to write in accord with communicative purposes were far from successful. What is just as important as the teacher's confused attempts to teach appropriateness, ending in highly inappropriate suggestions, is the emphasis she placed on the test rater. It seemed that the significance she attached to writing in an appropriate style was mainly due to the importance she attached to the rater. In other words, she appeared to be teaching the given aspect not mainly for the sake of achieving communicative purposes but rather for the students to impress the rater so that they would be rewarded with higher scores. Obviously, these findings highlight the importance of teacher factors such as teacher beliefs, training and competence in the final shape washback assumes. It seems that it is not easy for intended washback to magically pass through the teacher barrier. When teachers do not have a thorough understanding of the philosophy of a test or its demands, they are naturally unlikely to achieve the purposes of that test. Similarly, it seems that if the test demands are beyond what they can or have been trained to do, the test can hardly be successfully inflicted on teachers from the top to induce them to teach in the way desired by the test designers.

In her exhaustive doctoral thesis, Wang (2010) focused on the details of the role of teachers' beliefs, perceptions, attitudes and training in generating washback. In her study, she investigated whether the revised College English Test (CET), a national, high-stakes test, had resulted in the desired changes, particularly in teaching methodology, and examined the teacher-related factors which enabled or inhibited such change. She administered a questionnaire to 195 teachers and conducted group interviews with 30 in order to understand the teachers' perceptions of the new exam and their beliefs about teaching and learning. Finally, she conducted a case study with six of the teachers, which comprised classroom observations and follow-up interviews. The study yielded baffling results, revealing the mind-boggling complexity of teacher behavior and its influence on the generation of washback as well as of the washback phenomenon in general. According to the findings of the questionnaire, clearly, the majority of the teachers had a negative attitude to the new CET both in terms of its validity and its impact on teaching. However, interestingly enough, another majority

also believed the test had caused them to adopt a more language-use oriented teaching methodology and to focus more on meaning rather than the structural aspects of the language. The subtle contradiction between the teachers' responses is not only a sign of a lack of understanding of Communicative Language Teaching (CLT), which the CET purports to encourage, but also a foreshadowing of yet more inconsistencies to come as was confirmed by interview and observation data. The results from the classroom observations revealed that the teachers relied heavily on traditional teaching methods with a substantial amount of focus on language forms even in the case of three teachers whose teaching methodologies were sprinkled with communicative activities. The underlying reasons for this, uncovered through the follow-up interviews, appear to be the teachers' own low oral proficiency levels, their lack of a full grasp of communicative teaching methods as well as the teachers' own beliefs about language learning. However, considering that the test did not include a speaking section, it is not possible to make a definite judgment about whether the teachers skipped most of the speaking exercises in their coursebooks due to the exam or for other reasons. Similarly, given that the test did not prove to be as communicative as it was claimed to be and that it mainly aimed to reconcile traditional methods with communicative methods, it is not possible to know exactly why half of the teachers observed seemed to make an effort to use communicative activities despite their confused ideas and amongst the vastly traditional and teacher-centered methods that lay at the heart of their teaching while the other half seemed to cling on tight to their traditional methods. One explanation for the varying behaviors of the teachers in terms of using CLT in their teaching could be the continuous interaction and interrelation of teacher beliefs with the properties and the (intended) purposes of the test. Under this rationale, the well-intentioned attempts of some of the teachers to include CLT in their teaching might well be viewed as latent washback effects, and the vague and amorphous nature of these effects might be due to both the test itself and the teachers' gradually changing beliefs in response to the test, the progress of which may have been inhibited because of both the weaknesses of the test and a lack of training in CLT on the part of the teachers. The same logic could easily be applied to the teachers who showed a total lack of interest in adopting CLT methods, with the additional note that "tests will have washback effects for some learners and some teachers, but not for others" (Alderson & Wall, 1993, p. 121). However, it is important to note that although washback cannot be considered as a proof of test validity or lack thereof, with all due respect, it also seems to be difficult to

disentangle washback effects from other causes and reach a relatively definite conclusion in the absence of a test with strong construct validity and which is strongly aligned with the relevant curriculum.

Shohamy's report of the results of a number of studies about classroom assessment provides further interesting insight regarding the importance of teacher perceptions and training (1998). Shohamy (1993, as cited in Shohamy, 1998), Shohamy, Donitsa- Schmidt and Ferman, (1996, as cited in Shohamy, 1998) and Ferman (1998, as cited in Shohamy, 1998) investigated the influence of external tests on classroom assessment practices. They conducted surveys, interviews, observations and analyses of the teaching materials. Interestingly, yet not unpredictably, all of these studies found that external tests had substantial influence on classroom assessment. Teachers often administered assessments which mimicked the external test to be administered both in format and content. Moreover, it was observed that the degree of similarity between the classroom tests and the external tests intensified as the external tests drew near. Another common finding of these studies was that teachers without sufficient training in testing considered external tests as sound models and exploited them in an attempt to compensate for their sense of lack of knowledge of what a good test should be like. External tests thus played an informative role in classroom assessment by forming the basis on which teachers built their understanding of quality testing, which resulted in more preferable classroom assessment practices. Based on the notion that washback can be claimed to exist in the presence of things that teachers and learners do because of the test, but "would not necessarily otherwise do" and the fact that classroom assessment is a natural part of teaching and learning, it is possible to suggest that these findings show the external tests in question had a positive effect on classroom practices (Alderson & Wall, 1993, p. 117).

In addition to national tests, international tests such as TOEFL and IELTS have been the focus of a number of researchers. The pioneering example of such studies was carried out by Alderson and Hamp- Lyons (1996). The aim of their study was to investigate whether the-then paper-based TOEFL test induced negative washback on teaching and learning as it was widely perceived to do. For this purpose, they conducted student interviews, teacher interviews and classroom observations of two teachers while teaching TOEFL preparation classes and while teaching non- TOEFL classes at a well-reputed language institute in the USA. The student interviews suggested that the lessons were typically teacher-dominated and there was a lack of focus on the needs of

individual students. The findings of the teacher interviews showed that the teachers had negative attitudes toward the exam on the grounds that it was unnatural and not reflective of communicative competence. Also, they obviously felt under a lot of pressure, especially in terms of time. On the whole, from the self-accounts of the teachers, a typical picture emerges of a group of lethargic, jaded teachers who simply muddled along, where TOEFL preparation classes were concerned. The classroom observations did not fail to corroborate the accounts of either the students or the teachers. It transpired that there were vast differences between how the teachers taught TOEFL classes and how they taught non-TOEFL classes. The TOEFL classes were typically characterized by heavy teacher talk, little student participation and a very serious atmosphere in the classroom. The non-TOEFL classes of both teachers, however, were remarkably more student-oriented, included (a lot more) pairwork and there was a much lighter atmosphere in the classroom. These findings leave almost no room for doubt that the exam exerted important negative washback on teaching, especially considering the fact that both teachers displayed changes in their teaching practices in the same areas. However, although the change was noteworthy for both teachers, focusing on the fact that the teachers in question were not exactly affected to the very same degree, the researchers, yet again, bring up the question of the role of teachers' beliefs about teaching and their perceptions of a particular exam in the mediation of washback and the level of intensity it reaches.

Another TOEFL study, a more comprehensive one, which was conducted by Wall and Horàk (2011) provides a broad perspective on the question that arises from the Alderson and Hamp-Lyons's study (1996). The aim of their well-known, longitudinal study was to investigate the influence of the changes in the TOEFL exam on instruction in exam preparation classes in Central, Western and Eastern Europe. They started off with a baseline study with 10 teachers in 6 countries in 2003 before the plans to revise the test were officially announced in order to identify what teaching was normally like in the classes in question. Interview and observation data indicated that the teaching methodologies employed by the teachers were not representative of the recommended communicative language teaching practices. In terms of content, in line with the exam being used at the time, the computer-based TOEFL (TOEFL CB), the teachers obviously put a lot of emphasis on writing, vocabulary and reading as well as difficult grammar points; however, they did not focus on speaking as a separate skill as it was not included in the exam. Moreover, the teachers, who had not taken the exam

themselves except for one and with no special training for teaching TOEFL classes, relied heavily on their coursebooks to inform them about the characteristics of the exam. In fact, their teaching was mostly informed by the coursebooks themselves (Wall & Horàk, 2006) The second phase of the study, which was conducted with 6 teachers after the announcement of the new exam, the Internet-based TOEFL (TOEFL IBT), revealed the teachers' concerns over lack of sufficient information about the nature of the exam and lack of revised coursebooks although in general they had a positive attitude to the changes in the exam and expressed their willingness to implement changes in their teaching (Wall & Horàk, 2008). The results of phases 3 and 4, which covered the period just before and one year after the new exam was introduced in the countries in question, point to a strong influence of teaching materials on the teachers' teaching. As described by Wall and Horàk (2011), the teachers were heavily dependent on their coursebooks. Their anxiety over the new exam was slowly replaced by increased confidence as they got access to new coursebooks designed for the new exam, and they used the coursebooks as their main source of guidance, particularly in terms of exam content and format. However, in terms of teaching methodology, although some promising changes were observed, they varied in degree, from teacher to teacher and were less strong in comparison with the immediate and dramatic changes made to the teaching content. In accord with many other researchers, Wall and Horàk (2011) list training, experience and teachers' personal beliefs about teaching as the possible factors which may account for the varying degrees of change in the teachers' methodologies. However, one remarkable point about the report, deserving of particular attention, seems to be the emphasis placed upon the influence of the coursebooks on the teachers' teaching methods. It is clearly stated that the teachers experimented to varying degrees with different techniques based on the different coursebooks they used. Given the noteworthy relationship between the levels of training the teachers received and the changes they implemented, that cannot be asserted as standing in sharp contrast to Wall and Alderson's earlier conclusion (1993), which implies that coursebooks are unlikely to exert any influence on teaching methodology in the presence of insufficient teacher training. However, from a different perspective, it may also be considered as an indicator of the existence of a potentially two-way relationship between teacher training and the coursebook, that is, just as teachers will interpret the coursebook based on the level and quality of training they have received, the coursebooks they use may also

become a part of their training, albeit very slowly, and may thus slowly influence teaching and play a role in the generation of long-term washback.

Hayes and Read's IELTS study yielded similar results with regard to differences in teacher reactions to tests (2004). The aim of this widely cited study was to examine the washback effects of the Academic Module of the IELTS test on two different types of courses offered by two different institutions in New Zealand as well as its influence on course structure and design. In the first phase of their study, they found through survey responses that the majority of the schools they contacted offered courses specifically designed for the Academic Module, and the interviews with the teachers administered subsequently showed that teaching was mainly centered upon what the exam covered, with little attention to academic study skills. The focus of the second phase of the study was to investigate the washback effects of the test on two different IELTS preparation courses, one being strongly test-focused and the other with a broader focus on academic skills. Through classroom observations and weekly interviews with two teachers, they found that the test-focused course was dominated by the use of IELTS practice tests, the teaching of test-taking strategies and a strong focus on listening at the expense of any attention to the students' language knowledge and different skills whereas the other one incorporated different materials and focused on a broader range of skills, including those not tested on the test itself. Also, it was more student-centered and focused more on the general language abilities of the students. The pre and post- test results showed no significant change in the exam scores of the students taught by either teacher except for the improvement in the listening scores of the students who were taught in the test-focused class. Given these results, it is easy to conclude that the test exerted strong washback on teaching and learning in the test-focused class. Most of it was negative since teaching and learning revolved around the test with no attention to the language needs of the students. However, the improvement in the scores on listening tasks may also be considered as a positive effect based on the Washback Hypotheses of Alderson and Wall (1993), which includes the principle that "a test will influence the **degree** and **depth** of teaching; and a test will influence the **degree** and **depth** of learning" (p. 120). Despite this seemingly simple and obvious cause-and-effect relationship, the fact that the two courses displayed considerably different features compels further analysis of the situation, which naturally includes examining different factors other than the test itself such as the teachers and the educational settings. Such an examination quickly reveals the role of the teachers' own

perceptions in creating washback as they each focused more on the skills which they considered to be more important. What is maybe equally worthy of attention is the influence of the different attitudes of the two institutions toward the teachers and toward preparing students for the exam. The obvious differences between them in terms of course length, flexibility, teacher support and materials seem to have resulted in different washback effects. Not surprisingly, the shorter program, which offered no support to the teacher in terms of materials or course design became the test-focused course in the end, which was mostly negatively affected by the exam. The other program, which supported the teacher both in terms of materials and flexibility as well as course length did not seem to be significantly affected by the exam. All these factors and results, when considered as a whole, reveal yet again that washback effects are often dependent on different factors and can rarely be traced to a single cause.

Another prominent IELTS study was carried out by Green (2006). With a view to filling the gap in the literature with respect to the relationship between test design and those who are affected by the test, he investigated the influence of the writing component of IELTS on teaching with a special focus on the characteristics of the test. He conducted classroom observations of a total of 20 teachers and 197 students in 22 IELTS preparation and 13 EAP classes. The classroom observations of the twenty-two of these classes were followed by post-observation interviews carried out with some of the teachers. Also, a student questionnaire was used to identify the students' perceptions of the exam and student work was analyzed for further comparative purposes. The results of the classroom observations showed that the two types of courses shared quite a few similarities with respect to course structure, the types of topics covered in class, the use of certain instructional methods as well as the focus on the core aspects of formal written language. Both types of classes seemed to be characterized by whole-class interactions for the most part, which mainly focused on impersonal topics rather than personal or academic. Similarly, brainstorming activities, planning, writing practice and the teaching of a formal style of writing often featured in both types of courses. However, despite the overall similarities, the two courses showed notable differences in certain areas such as the amount of time devoted to the teaching of language form and the number of topics covered in each class. According to the frequency data, grammar and vocabulary took up a larger portion of class time in IELTS preparation classes. Also, IELTS preparation classes covered a larger number of topics in each class, thus treating the topics on a more superficial level in comparison with

EAP classes. More important than these was the divergence between the two types of courses with respect to the focus placed upon the fundamentals of academic writing which distinguish it from other styles of writing, in IELTS preparation classes, a complete lack thereof being the case. Whereas students were taught about plagiarism, hedging and to reference their sources in EAP classes in addition to the overall focus on the features of formal language, IELTS classes were observed to focus only on vocabulary and structures as well as formulaic phrases that might help students on the test with no attention whatsoever to the requisite writing skills essential to producing higher-level academic texts that the students are expected to write in higher education contexts. An extension of this difference could be seen in the approaches and materials adopted by the teachers with a view to helping students practice and deepen their understanding of what they had been taught. In IELTS classes, all of the textbooks used were exam preparation books, and the intense focus on test content was accompanied by timed writing activities whereas in EAP classes, the materials used did not include test-preparation books and students were asked to conduct research and do projects relevant to their academic subjects, neither of which activities was encouraged in any of the IELTS preparation classes observed. Predictably, student work reflected these practices: The texts written by the IELTS preparation students consisted of 98 to 445 word- essays whereas the length of those produced by the EAP students varied between 128 and 3,495 words. Maybe one of the most significant indicators of intense focus on the test was the explicit mention of the test by the teachers in IELTS preparation classes. The test was mentioned 129 times during those classes and only 10 times during EAP classes, nine of which took place in classes on combination courses. The interview data showed that the teachers teaching the IELTS Preparation courses considered IELTS as an aim of the class and the objectives they listed were directly, and some of them relatively implicitly, related to test tasks and demands. Although similar objectives were mentioned for the EAP courses, the teachers also mentioned several other aims and objectives, which did not seem to be shared by IELTS preparation courses. Given these findings, it is possible to say that the test exerted some negative washback effects on teaching as it seems to have resulted in an intense focus on test demands, which are directly linked to test design, at the expense of excluding the teaching of certain essential academic writing skills. However, from another perspective, the test may be considered to have had positive effects considering the characteristics the IELTS preparation courses had in common with the EAP courses, however, only if we assume

that the teaching practices observed in the EAP classes displayed the typical features of ideal teaching practices. Regardless, it is not possible to reach a definite conclusion without a deeper analysis of teacher factors such as teachers' beliefs and teacher training as these factors might have played a role in the teaching practices in question (Green, 2006).

Badger and Yan (2012) investigated the influence of IELTS on teaching in the context of China. The main interest of their study was to examine how much of classroom instruction was characterized by CLT in IELTS preparation classes. In order to uncover any potential contextual factors, they administered a questionnaire to 69 teachers, which addressed the different aspects of the educational settings they worked in as well as their approaches to teaching and learning. Subsequently, the researchers conducted interviews with 10 teachers to support the questionnaire data. Finally, three different teachers were observed while teaching IELTS classes, followed by post-observation interviews. According to the results of the questionnaire, administrators had a considerable amount of control over the choice of materials, which mainly consisted of IELTS preparation textbooks. As for the teachers' approaches to teaching, it seemed that the teachers mainly preferred to use methods associated with CLT such as using English as the medium of instruction, focusing on language functions rather than form and utilizing open-ended questions to encourage extended speech. Interestingly enough, it was also found that the teacher was considered as the main locus of control. Most of the findings from the questionnaire data were confirmed by the interview data. Moreover, the interviews with the teachers also revealed some interesting details which could not have been elicited by the questionnaire alone. One of them relates to the teachers' perceptions of the role of the teacher. It seemed that although the teachers mainly adopted a teacher-centered approach to teaching, many teachers also expressed that they assumed different roles in the classroom ranging from a controller to a facilitator and a guide, and one teacher pointed to student expectations as an important reason for the adoption of a teacher-centered approach. Another important detail the interviews found was that although the teachers held that teacher talk in English provided their students with an opportunity to improve their speaking and listening skills, the same teachers preferred to use L1 while carrying out the more technical aspects of teaching such as the teaching of skills and text analysis. It seemed that the teachers took a different attitude to speaking English in class while teaching about the language and when encouraging their students to practice their language skills. The

results of classroom observations were mostly consistent with the findings of the questionnaire and the interviews with respect to the teachers' approaches to teaching. The teachers seemed to lean toward CLT in terms of their teaching styles, which was often vitiated by the excessive use of L1 during the classes. It also appeared that not all of the teachers' classes were characterized by teacher domination. In line with the findings from the interviews, the results of the post-observation interviews point to the role of the differing beliefs of the teachers regarding teaching and learning, contextual factors such as time constraints and student expectations as well as the training the teachers themselves received in influencing the teachers' decision to resort to L1 and to use different methods in their teaching. Based on these findings, it is possible to suggest that IELTS exerted obvious washback effects on teaching content, whereas its influence on pedagogical practices seems to have been rather vague and indirect. Although the teachers' adoption of a communicative approach to teaching could be linked to their perceptions of the test, how they implemented this approach seems to be more related to their beliefs and contextual factors than to the test itself, which supports the findings of previous research.

Bearing in mind the evidence provided in the literature with regard to the role of the teacher in the generation of washback, Chappell, Bodis and Jackson (2015) specifically aimed to explore the influence of teachers' perceptions and understanding of IELTS on their instructional practices in exam preparation classes in the context of Australia. For this purpose, in the first stage of their study, they administered a questionnaire to 40 teachers in order to gauge their attitudes and understanding of the IELTS test. Following the administration of the questionnaire, interviews were conducted with 10 teachers to gain a deeper understanding of their perceptions. Finally, each of the teachers who had taken part in the interview was observed for 1-2 class hours in order to identify the teachers' approaches to teaching IELTS classes and to compare the teachers' stated beliefs with their actual classroom practices. The findings revealed that most of the teachers lacked a thorough understanding of the test and had a negative attitude toward the test for one reason or another. The classroom observations showed that each teacher taught IELTS classes in a different way based on her/his own personal perceptions of language learning and/or her/his notions of the exam. Points of convergence and divergence between the teachers' stated beliefs about teaching and learning and their actual classroom practices were also identified through the observations. It is important to note that the inconsistencies revealed in this area do not

suffice to make a claim about the unreliability of teacher accounts as most of these inconsistencies seem to have stemmed from their perceptions of the exam. Thus, it may tentatively be suggested that these inconsistencies actually point to washback and in this case negative, teacher-mediated washback effects according to the researchers, as these inconsistencies resulted in the courses in question being reduced to teaching and learning test-taking strategies and to General English classes with no sufficient focus on the specific requirements of the test itself.

The growing interest in improving assessment practices has resulted in an upsurge of washback research. Nevertheless, the number of observational studies remains small in proportion to non-observational studies. The gap seems to be distinctly wider in classroom assessment research. However, what is even rarer is research on the influence of classroom assessment on teachers and teaching. One has to dig very deep into the existing literature to get her hands on studies which combine both features.

One of such rare studies belongs to Burrows (2004). As part of her doctoral research, she investigated the washback effects of the-then newly-introduced Certificate of Spoken and Written English (CSWE) for the assessment of the English language competence of new adult immigrants to Australia. CSWE was a type of classroom-based assessment, induced by the state but applied and administered by teachers inside the classroom. Teachers were explicitly guided through the new system via the explicit specification of learning outcomes and provision of example performance tasks as well as specific criteria against which to assess student learning. Burrows (2004) conducted a survey of 215 teachers, interviews with 30 teachers and classroom observations of four teachers in order to examine how this externally controlled, but internally administered assessment affected classroom practices. Both the survey and the interview results showed that many teachers believed the new assessment had resulted in changes to their teaching. The classroom observations of the teachers, selected based on the interviews, mostly supported the teachers' perceptions of the change or lack thereof that had occurred to their teaching. Drawing on the observation data, Burrows (2004) identified four groups of teachers in terms of their reactions to change: resisters, adopters, partial adopters and adaptors. She attributed the differences between the teachers' responses to change mainly to an element of choice, pointing out that "there is a degree of choice involved in washback, at least in the context of classroom-based assessment" (p. 125). Since the choices people make are often closely linked to their understanding and perception of a given situation, Burrows's findings support the evidence in the literature

that washback does not exist in a vacuum independently of teachers, who are probably the most important mediators of washback. Teachers' pedagogical beliefs and their perceptions of an exam often have a direct bearing on the type of washback that occurs in the classroom (Hughes, 1993, as cited in Bailey, 1996; Burrows, 2004; Wall & Horàk, 2011).

Olovsson's study yielded much more inauspicious results (2015). The focus of his research was the relationship between learning goals, teaching and assessment. In 2011, he set out to explore what assessment looked like in the classroom, following the introduction of a series of reforms by the government in order to increase success rates in Swedish schools. These reforms involved the implementation of a set of changes in the national curriculum and syllabi and laid greater emphasis on grade reporting for the purposes of tracking student success and increasing accountability. For his study, Olovsson conducted 40 hours of classroom observation as well as teacher and student interviews. The observation results showed that the specification of learning goals, which was core to the reforms, consisted of weekly plans designed by the teachers themselves. These plans did not extend beyond informing students about the tasks to be completed and were mostly exclusive of any detailed explanations of the standards students were expected to reach. Simply put, the weekly plans were more like weekly segments of the national syllabi and since the tasks included in those plans had to be completed within a certain time frame, time constraints also seemed to constitute a major problem. Thus, highly unsurprisingly, both teaching and learning were strictly focused on the completion of a series of tasks rather than the development or improvement of the relevant skills. The teachers encouraged the students to produce quick, superficial answers. Also, most of the activities were limited to gap-filling exercises and producing short answers. As for feedback, the students received mostly evaluative feedback rather than descriptive. In other words, teaching and learning were mostly reduced to keeping up with the appearances: the tasks were completed, however at the expense of internalized learning and higher-order thinking skills. The interview data supported the findings from the observations. Neither the teacher nor the students interviewed had a clear understanding of the expected learning outcomes. Their perceptions of learning goals were limited to the relatively narrow content of the weekly plans, and both parties attached great importance to the completion of tasks. The teacher mentioned that the increased emphasis on grade reporting had resulted in increased pressure on teachers in general. This concern seems to be closely related to the

teacher's, and thus the students', self-reported accentuated focus on achieving good grades by completing the relevant tasks without much regard to the actual learning goals specified in the national syllabi. However, the same teacher also stated that their main goal was to help the students achieve the necessary grades so that they could move up to a higher-level of education. This statement is important in that it could easily be taken as an indicator of her own perception of the role of either the assessment system in question or assessment in general in the teaching/learning process rather than that of a direct influence of the new assessment system. Thus, although it stares us right in the face that the new system apparently had very few if any positive effects on the teachers and students observed, it is important to note that the teachers seemed to have a vague idea of the real nature of the learning goals. What is equally important to bear in mind is that the aforementioned statement expressed by one of the teachers also suggests that the teacher's own perceptions and beliefs may have stood in the way of any potential positive washback. In short, the findings of Olovsson's study are consistent with those of Burrows's and numerous other studies and provide yet another piece of evidence that turning a blind eye to teacher perceptions and expecting assessment to reform teaching altogether on its own hardly goes beyond being wishful thinking.

Luckily, there are also studies which show it is not all doom and gloom. One of them was conducted by Bahat et al (1997, as cited in Shohamy 1998). The purpose of their study was to examine the effects of a pilot project implemented in 22 schools in Israel on classroom practices, teachers' and students' attitudes toward assessment and their perceptions of language ability. The project involved the use of alternative assessment procedures in place of the regular large-scale, national end-of-high school examination. The teachers received intensive training and supervision from experts to implement such assessment in the classroom. Bahat et al (1997, as cited in Shohamy 1998) included one of these schools in their study. They conducted observations as well as surveys and interviews with teachers, students and principals. The findings suggested very promising outcomes. The teachers included more creative and engaging tasks both in their teaching and assessment practices in comparison with the regular classes. Also, both the teachers and the students seemed to have a positive attitude toward the new assessment system. Rather than considering it as an end product, they now perceived assessment as an integral part of the learning process. A similar positive attitude was observed in the understanding of language ability. The perception of language knowledge had moved beyond the internalization of what was tested on the regular

national test as the standard. Obviously, the absence of an external test had resulted in considerable positive changes in teaching and learning (Shohamy, 1998). The key factor that led to these uplifting results could have been the practical intensive training the teachers received, which may have helped the teachers change their beliefs while also guiding them through the new assessment procedures, thereby reforming teaching.

Munoz and Alvarez's study provides further support for the propitious potential of joining hands with teachers to promote positive washback (2010). The purpose of their study was to investigate the washback effects of a new oral assessment system on teaching and learning at a language center in a university in Colombia. The Oral Assessment System (OAS) was a classroom- based assessment system designed with particular attention to the principles of communicative language testing, and thus, it was expected to create positive changes in the teaching and learning of oral skills. The researchers followed a comparative study design in order to explore the influence of teacher training on the generation of washback, and their data collection methods consisted of a student survey, a teacher survey, 54 hours of classroom observation and external evaluations of student success. The results revealed notable differences between the classroom practices of the teachers in the experimental group, who received ongoing support in addition to training on how to implement the rubrics in accord with communicative language testing and those of the teachers in the comparison group, who had not been provided with any supplementary guidance. According to survey data, all the teachers in the experimental group stated lesson objectives in written form. Also, sixty per cent of those teachers reported stating the objectives orally. As for the comparison group, half of the teachers stated that they specified objectives orally. However, the difference suggested by these results pales into insignificance in comparison to the startling results of the classroom observations. The observations revealed that whereas the vast majority of the teachers in the experimental group specified the objectives both orally and in written form, put plainly, none of the teachers in the comparison group stated any specific objectives. Also, whereas the majority of the teachers in the experimental group used instructional tasks that were relevant to class objectives, the same was true for only a small proportion of the teachers in the comparison group. Bearing especially that in mind, one would not feel dumbfounded by the finding that the teachers also differed in the way they taught. Although both groups of teachers stated that they used a variety of activities in their teaching, the classroom observations confirmed this claim only for the teachers in the experimental group. It

was revealed that the teaching of the teachers in the comparison group mainly consisted of traditional methods and activities focused on language forms. Other significant differences found between the two groups are related to the assessment tasks used by the teachers and the ways they reported assessment scores. It was found that the teachers in the experimental group mainly used assessment tasks that were similar to the tasks used in class. However, the use of completely new tasks was the most common practice among the teachers in the comparison group. The main difference between the two groups with respect to score reporting was the approaches they adopted to evaluate student performance. Whereas the teachers in the experimental group took an analytic approach, focusing on the different aspects of the language, the teachers in the comparison group tended to have a holistic approach. The results of the student survey reflected the teachers' approaches. The students in the experimental group had a better understanding of the assessment criteria specified in the rubrics. Also, the use of the rubrics for self-assessment was significantly more common among the students in the experimental group. Finally, although the external evaluations did not reveal a great difference between the two groups in terms of task completion and vocabulary knowledge, they showed that the students in the experimental group were more successful in communicative effectiveness, pronunciation and grammar. These results show that supporting teachers and standing beside them during the adaptation process might be more beneficial than simply pushing them to carry out the task of assessment that is set before them. It seems that when teachers are provided with the necessary training and support, they are less likely to be resistant to change and their classroom behavior might thus become more conducive to the generation of positive washback.

2.7.4. Washback Studies in Turkey

Despite growing interest in recent years, washback remains a new area of research in Turkey. As a matter of fact, Turkish washback literature consists of a limited number of studies, most of which rely heavily on questionnaires and interviews. Currently, there are remarkably few Turkish studies available in the literature which involve direct observation of classroom practices. Unfortunately, it appears that classroom observation has yet to be established as a standard component of washback research in Turkey. Thus, the relevant literature produced in Turkey provides little information with respect to the influence of tests on actual classroom practices despite

offering valuable insights into the potential impact of exams on the perceptions and attitudes of test participants.

One reason for this lack of observational studies could be that in contrast to the general trend in washback research around the world, most research carried out in this area by Turkish researchers in Turkey has focused on the learner's perspective rather than the teacher and teaching practices. One of the most well-known examples of such studies available in the Turkish washback literature was conducted by Özmen (2011). The purpose of his study was to investigate the washback effects of the Inter-university Foreign Language Examination (ILE) on test-takers. At the time of the study, ILE was a gate-keeping test which was aimed at candidates wishing to pursue an academic career or a graduate degree at a university in Turkey. Also, candidates for promotion to Associate Professor were required to take the exam to provide evidence of their foreign language ability. In terms of its design and content, the test was completely out of line with the principles of communicative language testing. It was essentially a multiple-choice test of grammar, vocabulary knowledge and reading comprehension, which did not cover listening, speaking or writing skills. In order to explore how test-takers were influenced by this highly traditional, very high-stakes test, the researcher conducted interviews with 12 prospective test-takers, all of whom were enrolled in a private preparation course. Considering the nature of the test, the results were not shocking. The participants' accounts simply mirrored the design of the test. All of the students interviewed bemoaned the fact that because of the test, they had to focus too much on grammar and vocabulary. All they did was to study grammar items and memorize vocabulary. They did not believe ILE was an effective exam which reflected one's actual language ability. Also, it was obvious that some of the participants deeply begrudged the time they had to devote to exam preparation. They considered the process as an unnecessary obstacle in their way, as a waste of their precious time and money. A similarly grim picture emerged when they were asked about the influence of ILE on their attitudes to foreign language learning. It seemed that the participants had grown weary of the process and had little hope of improving their communicative skills as a result of their preparation for the exam. Some of them reported that ILE had dampened their enjoyment of language learning whereas some had come to view language learning as a flat-out uninteresting and unpleasant experience. From these results, it is evident that the exam had a negative effect on the participants' attitudes. Their experience of language learning was reduced to learning *about* the language and

the exam, leading to feelings of frustration and decreased motivation. A more interesting finding of the study, however, was the fact that the younger participants were less resentful of the exam and viewed it in a more positive light than the practicing academics, which shows that student age and motivation are among the factors that influence the generation of washback.

The same test has also been the focus of two Master's theses. One of the relevant studies was carried out by Dağtan (2012). In order to explore the washback effects of the test in question, he conducted a survey and interviews with academics working at a state university in Turkey. The results were consistent with the findings of Özmen's study (2011) in that vocabulary and grammar were the two language aspects that the participants made the most effort to improve as they considered the relevant questions on the test to be among the most challenging. Another finding common to both studies concerns the appropriateness of the test in assessing foreign language competence. As in Özmen's study (2011), the majority of the participants did not perceive the test as an accurate assessment of language ability mainly on the grounds that it did not cover all language skills. Although the overall attitude of the participants does not evoke the same gloomy picture of a despondent group of test-takers that emerged from Özmen's findings (2011), they also complained that the test did not promote the learning of communicative language skills that would better serve their needs in their academic lives. The only glimmer of a positive effect exerted by the test as found by the study seems to be the participants' relatively favorable impression of the examination with regard to its potential to help improve reading skills.

The other study was carried out by Çakıldere (2013). For his research, which also included the investigation of the effects of a kindred test, he conducted a survey and subsequently interviews with academics working at a state university in Turkey. The results were painfully redolent of those of the two aforementioned studies. It was found that the vast majority of the participants focused greatly on grammar, although they deemed grammar to be the least important aspect of language learning. They also attached great importance to vocabulary and reading due to the structure of the test at the expense of all other language skills. Based on the understanding of washback as the things teachers and learners do because of the test, but "would not necessarily otherwise do," it is possible to interpret this intense focus as positive washback (Alderson & Wall, 1993; p. 117). However, at the same time, the fact that the participants relied heavily on test-taking strategies which consisted of memorizing some rules and grammar structures

casts a shadow over the "positivity" of this effect and renders such an interpretation highly debatable. With regard to perceptions of the test, the interview results revealed that just like the participants in the other two studies, the participants were unanimously of the opinion that the test could not accurately assess language ability as it did not cover all of the four language skills. It was also found that the test caused stress and anxiety for most of the participants, which had a negative influence on their attitudes toward learning English. Obviously, these results unequivocally confirm the findings from the two aforementioned studies that the test in question was highly likely to exert remarkably strong effects on the test- takers' attitudes and behaviors, most of which seem to be on the unfavorable side.

Yıldırım (2010) also focused on a very high-stakes national test. However, instead of examining the pre-examination process, he took a retrospective approach that invited college students to share their experiences regarding exam preparation. The test he chose to focus on was the English component of the national matriculation test, which students are required to take as part of the examination in order to be admitted into an English-related undergraduate degree program. The participants consisted of 70 college students enrolled in the English preparatory year program of the English Language Teaching Department of a state university in Turkey and 6 English instructors working at the same university. In order to collect data, he administered a questionnaire to the 70 students. Subsequently, he conducted interviews with 10 of them and finally with the teachers. Although the test is different from the previously mentioned one in terms of its purpose, due to the striking similarities between them in terms of their format and structure, it is by no means surprising that it exerted similar effects. It was found that according to the experience of the student participants, during their senior year in high school, instruction was mostly delivered in Turkish and the lessons were remarkably dominated by grammar, vocabulary and mock exams. Also, it seemed that the teaching and learning activities mostly consisted of teacher explanations about grammar, memorizing vocabulary, doing reading exercises and discussing various test-taking strategies although a small minority of the students reported doing speaking, listening and writing activities as well. As a result, although the students believed exam preparation lessons had helped them improve their grammar, vocabulary and reading, which they found helpful at college, they did not believe the same was true for their writing, listening and speaking skills. Most of the participants did not believe performing well on the exam was proof of one's ability to communicate in English. For

the most part, the instructors' analyses were in line with the students' perceptions and opinions. They stated that students' overall language abilities were negatively affected by the test as it did not cover all language skills. They pointed out that the multiple-choice format was another problematic feature of the test as it did not encourage the students to focus on producing the language or thinking critically, as a result of which they had difficulty in their studies at college. These results explicitly show that the test had negative washback effects on the students' productive skills. However, what is less emphasized is the fact that apparently, a few teachers persisted and insisted on teaching listening, speaking and writing skills although they were by no means assessed on the test. That confirms the hypothesis that "tests will have washback effects for some learners and some teachers, but not for others" (Alderson & Wall, 1993, p. 121). However, due to the complete lack of data regarding the teachers who actually taught these students during the exam preparation process, it is impossible to know exactly what factors caused them to act differently from many others. (Watanabe, 2004a).

Kılıçkaya (2016) directed his attention to primary education. He examined the washback effects of the English test included in the Transition Examination from Primary to Secondary Education (TEOG), another high-stakes, national test, administered to eighth- grade students seeking admission to a selective high school in Turkey. In addition to being the first study to investigate the effects of the English component of the examination in question, Kılıçkaya's study is also one of the few Turkish washback studies that have foregrounded the teacher's perspective. He aimed to investigate the effects of TEOG on the teaching practices of language teachers, and for this purpose, he conducted semi-structured interviews with 30 teachers teaching eighth-grade students in different schools. The results showed the teaching practices of the participants mostly mirrored the content and format of the test. Although all the teachers showed an understanding of the importance of communicative language skills, almost all of them reported focusing their instruction on grammar, vocabulary and reading. Except for two participants, none of them included speaking, listening or writing skills in their teaching as they were not tested on the exam. Their teaching style was strongly teacher-centered and was characterized by traditional methods such as grammar translation and lecturing. In terms of materials, they used supplementary materials such as test preparation books or exam-oriented coursebooks to better prepare their students for the exam. The assessment procedures they used were limited to tests and homework, most of which consisted of multiple-choice questions and gap-filling exercises.

Considering that all of the participants had an understanding of the importance of communicative skills, these results suggest that the test had negative washback effects on their teaching practices. However, it is worth noting that knowing about the importance of communicative skills does not necessarily mean being able to implement the necessary teaching methods. What is even more important is that without an analysis of the teaching methods that the participants think are normally ideal and would prefer to use were it not for the test, it is not easy to sort out the actual washback effects from teacher factors, which could easily amplify such effects. Thus, although it is evident from these results that the test led the teachers to act in certain ways, it is not clear whether the examination caused the teachers to adopt the teaching methods that they did or whether it actually simply reinforced the use of some such methods.

Although Turkish washback literature differs remarkably from the world literature in that it is characterized by a stronger focus on the student than on the teacher, with regard to the interest in national tests, most Turkish washback studies parallel many other studies carried out in different parts of the world. As is the case around the world, large-scale, national tests have received more attention from Turkish washback researchers in comparison to classroom-based assessment. One of the few Turkish studies that have focused on the influence of classroom-based assessment was conducted by Duran (2011). She aimed to explore teachers' and students' perceptions of classroom-based speaking tests and the washback effects of such tests. The participants in her study were 307 students following the English preparatory year program offered at a state university in Turkey and 45 English instructors teaching in the same department at the same university. In order to collect data, she administered questionnaires to the teachers and the students. Following that, she conducted interviews with 7 of the students and 6 of the teachers. The results revealed a very positive attitude toward learning and teaching speaking skills. The majority of the participants agreed that they would be interested in teaching and learning speaking skills even if they were not tested. It was also found that the vast majority of the teachers believed the tests in question had a positive effect on their students' speaking skills, which was confirmed by the majority of the student respondents. However, there was considerable variability among the participants' perceptions with regard to the ability of the given tests to assess speaking ability accurately. This suggests that the participants did not have full confidence in the scores gained on the tests as a reliable indicator of student ability. Three important interview findings which may explain this

lack of certainty relate to problems with test content, deficiencies in the scoring system and test anxiety. Nevertheless, apparently, these flaws did not dampen the participants' positive view of the tests with regard to their effects on speaking skills. In terms of the influence of the tests on teaching practices, it was found that the percentage of teachers who claimed not to make any changes to their teaching because of the tests was almost equal to that of teachers who reported adapting their lessons according to the tests in terms of content and time allocation for speaking activities. This provides further support for Alderson and Wall's hypothesis that "tests will have washback effects for some learners and some teachers, but not for others" (1993, p. 121). Considering the teachers' claim that they often held extra speaking activities that were different from the ones on the test, this finding is also consistent with Burrows's argument that "there is a degree of choice involved in washback, at least in the context of classroom-based assessment" (2004, p. 125). The percentage of students who seemed to focus on exam content was slightly higher than that of the teachers, which suggests different groups of test participants may be affected differently by a test. Another important finding related to learning practices was that the teachers believed the tests encouraged students to speak more in class. However, some of the students' accounts suggest that students, as a matter of fact, at least some of them, developed a nonchalant attitude toward the tests as a result of repeated exposure to the same types of tasks and the fact that the tests were weighted at a low percentage, which raises doubt about whether there is as direct a relationship as it seems at first sight between the tests and student motivation. Nevertheless, taken together, these results show that the tests had strong positive washback on students' speaking ability whereas they do not seem to have affected the teachers to a very great degree. Given the lack of confidence in the structure of the test, this finding seems to confirm the argument that a good test may generate negative washback and a bad test may result in positive washback depending on the test participants' perceptions and what they do (Alderson & Wall, 1993; Hughes, 1993; Messick, 1996). In this study, it is clear that the participants had a very positive attitude toward the tested skill, which may account for their positive perceptions of the tests in spite of the reported problems. As a matter of fact, based on these results, it is possible to conclude that the tests generated positive washback by their sheer presence thanks to the attitudes of the test participants.

Köktürk (2015) examined the impact of the quizzes and the interim tests administered at longer intervals as part of the institutional examination system used at a

university. She conducted a comprehensive survey of 464 students, which was followed by interviews with 10 of the students and 6 instructors working at the same university. It was found that the vast majority of the students believed the quizzes were strongly consistent with classroom practices in terms of content. More than half of them were also happy with the frequency of the quizzes as they thought it kept them alert and prompted them to study to have frequent exams. However, although the majority of the students found it helpful to receive feedback about the correct and incorrect answers after the exams and believed the quizzes helped them learn the topics that they did not know or misconceived, interestingly, only a small portion of them reported that the quizzes motivated them to learn English. The interview data suggest that this could be related to the format and content of the tests, which required a great deal of memorization of grammar and vocabulary items. Also, it was mentioned both by the teachers and the students that the students did not take the quizzes seriously as they did not have a great impact on the final course grade. Thus, another reason why the quizzes failed to spur student motivation might have been the low weight assigned to them. The findings concerning the participants' perceptions of the interim tests testify to the important role that test structure and exam weightings can play in the generation of washback. Unlike the quizzes, the interim tests were reported to assess all four language skills. Apparently, these tests did not involve the direct and explicit testing of grammar and vocabulary. Instead, they required the students to demonstrate their abilities to comprehend and use the language in various ways, hence no particular need for memorization. Also, they were weighted more heavily than the quizzes. As a result, although the participants seemed to be uncomfortable about the unpredictability of the interim tests in terms of test content, the percentage of students who felt motivated by the interim tests to learn English was higher than that of those who were motivated by the quizzes. The teachers, however, seemed to be less inspired by the interim tests in comparison with the students. They stated that they felt unmotivated to teach grammar and vocabulary items due to a lack of student interest. As the students knew the interim tests did not test grammar and vocabulary explicitly, they tended to show no interest in learning such items, presumably during the lessons right before the examinations. Thus, although the teachers believed it was necessary to teach grammar and vocabulary, they felt as if they were wasting their time by doing so. The teachers also complained that the quizzes were given less weight, which inclined the students to disregard them and the topics tested on these exams in favor of focusing on the interim tests. Thus, it seems that

the exams had negative washback to some extent on the participants' interest in teaching and learning grammar and vocabulary items. However, the teachers' accounts indicate that speaking and writing tasks that students were asked to perform on the interim tests were the same tasks that were covered in class, which vaguely suggests that the tests may have had positive washback on teaching content. However, it could also be a reflection of the teachers' beliefs or simply the result of the system, or maybe a combination of all these factors. Much more mysterious is the classroom practices of the participants, as based on these results, there is no knowing how and to what extent the teachers and students practiced those tasks. One very valuable and definite conclusion that may be reached from results, however, is that giving students too many tests and testing them on too frequent a basis may diminish the significance of such tests in the eyes of the students and thereby give rise to the generation of negative washback, especially if they are also inauthentic and weighted lightly.

Despite their valuable contributions to the literature, the Turkish studies cited above provide little insight into teacher behavior in response to exams. The most fundamental reason for this seems to be the lack of direct observation of classroom practices. All of these studies are based only on teacher and student self-reports, which precludes verification of the influence of exams on actual practices as what teachers say they do and what they actually do can often be different (Alderson & Wall, 1993; Farrell, 2015; Lankshear & Knobel, 2004; Stiggens & Conklin, 1992). As of yet, the number of Turkish washback studies that have employed classroom observation is so painfully small as not to require all the fingers on one hand to be counted.

One of the earliest examples of washback studies in the context of Turkey in line with modern washback research around the world was carried out by Sevimli (2007). For her study, she conducted questionnaires, interviews and classroom observations. She aimed to examine the effects of the Foreign Language Examination (FLA) on secondary education, which students are required to take as a part of the university entrance examination in order to be admitted into an English-related undergraduate degree program at a university in Turkey. The results demonstrated that this test of critical importance had a tremendous impact on both the teachers and learners. In keeping with the requirements of the test, teachers merely taught grammar, vocabulary and reading through mock exams. They disregarded the other communicative skills as they were not included in the exam, which the researcher observed to lead to rote-learning. It was also found out that the test caused great anxiety in students. Based on

these results, it is possible to assert that the test had direct and obvious negative washback effects on what the teachers taught and how the students learned. Considering the fact that these teachers were under great pressure from all sides to enable their students to get into university and the time constraints they had to work within, it stands to reason that they did not focus on the teaching of listening, speaking and writing skills, which were not included in the exam. However, because of a lack of information about the teachers' pedagogical beliefs, it remains a nagging question how differently these teachers would teach in the absence of the exam and whether they themselves played a role in the intensification of the washback from the exam through their own beliefs about teaching.

Şentürk (2013) investigated the washback effects of the Key English Test (KET). For her study, she videotaped 8 KET preparation lessons and 6 General English lessons taught by the same teacher at a private secondary school in Turkey. In addition to that, she conducted interviews with 5 teachers working at the same school and administered a questionnaire to 20 of the students preparing for the exam. Analysis of the data from the videotape records revealed substantial differences between the teacher's exam preparation classes and General English classes. The exam preparation classes were characterized by a serious, tense atmosphere and teacher domination. The teacher was strict and employed only teacher-centered methods in her teaching. Classroom activities were limited to examination practice through an exam preparation book and test booklets from previous years. In the General English classes, however, not only was there a much more relaxed atmosphere and room for laughter and fun, but also the teacher typically used a variety of communicative teaching techniques such as pair-work, role-playing and dramatization. Teaching materials were also more varied and included a workbook, teacher-made worksheets and vocabulary lists in addition to a coursebook. Thus, it is evident from these results that the test had a negative impact on teaching practices by causing the teacher to simply teach to the test at the expense of instructional practices that encourage learning by doing in an authentic context.

More recently, Sağlam (2016) examined the influence of an institutional proficiency test on teaching practices in the preparatory English language program (PEP) of a private university along with whether the instruction program led to any gains in students' test scores and the appropriateness of the decisions based on students' test scores. Participants in her study consisted of students enrolled in the PEP, language teachers teaching PEP students, freshman students and instructors of mainstream

courses at the faculties. She gathered data through surveys, interviews, classroom observations and the pre- and post-proficiency exam scores of PEP students. The results revealed a close match between the content of teaching and the content of the test. In addition to a regular coursebook, institution-produced supplementary materials were frequently used by all of the teachers. As a matter of fact, the supplementary materials received greater focus and were used more frequently than the coursebook. One contributing reason for this intense focus on the supplementary materials may have been that the participants perceived them as useful and helpful in learning English. However, it is evident from the results that the predominant reason was the relevance of those materials to the test. Obviously, they were particularly geared toward the exam and functioned as exam preparation materials. As a result, both groups of participants, especially the students, considered the coursebook irrelevant due to a lack of similarity between the task types covered in the coursebook and those on the test whereas they paid particular attention to the supplementary materials. Data from the classroom observations confirmed the importance attached to those materials by the participants. In all of the classes observed, they were used as the primary teaching and learning resources. Most of the teachers strictly adhered to the content and format of the materials. As a consequence, although they seemed to adopt a skills-based approach, speaking was never the main focus of the lessons, obviously since the test did not assess speaking skills. Also, no activity leading to a meaningful communicative situation could be observed, which conflicted with the questionnaire results suggesting that teaching and learning practices often included communicative activities. With respect to gains in student test scores, the study found that the students had benefitted significantly from the instruction program. However, in terms of the appropriateness of the decisions based on these scores, it was revealed that the instructors did not consider students' test scores to be accurate indicators of their language abilities. More precisely, they believed many freshman students had considerable difficulty in understanding and using academic English despite having passed the proficiency test, which was out of line with the students' optimistic and positive views of their own progress. These findings show that the test in question exerted strong washback on the content of teaching and learning. Although the focus on three language skills via the exam-oriented supplementary materials could be considered as a positive washback effect, it is not possible to overlook the negative washback of the test on the practice of speaking skills. Apparently, both the majority of the teachers and learners clearly prioritized the so-

called supplementary materials over the coursebook and other materials, which resulted in a lack of focus on speaking skills. The test also induced negative washback on teaching methodology through the materials. The rigid adherence to the supplementary materials, which mimicked not only the content but also the format of the exam, inevitably caused teaching to be geared toward not only the content but also the format of the exam. In line with the multiple-choice format of the test, the majority of the teachers mostly elicited one or two-word responses from the students, and in contrast to what they reported on the questionnaires, they did not employ any activities that resulted in a meaningful communicative interaction although some teachers utilized a variety of sources in addition to the institution-produced supplementary materials and used open-ended questions in their teaching as well as teaching critical thinking and skills strategies. Bearing in mind the significant gains in students' test scores at the end of the program, it is not difficult to guess why most of the teachers simply relied on test-oriented materials and opted to teach to the test. However, based on these data, it is not as easy to understand why and how some teachers were relatively resistant to the negative washback effects of the test in comparison with the other teachers.

Apparently, another element missing from the Turkish washback research tradition is the examination of teacher beliefs. Considering that washback is a relatively new area of research in Turkey, this is not surprising because there seems to be a global shortage of washback studies which involve a systematic analysis of teacher beliefs. However, it is clear that not even direct observation suffices to explain why teachers differ in the way they respond to a given exam and thus, it seems that unless the role of teacher beliefs in the generation of washback is established, any interpretation made on why teachers do what they do in response to an exam will remain in the realm of educated speculation.

2.8. Personal Construct Theory

Personal construct theory is a psychological theory of personality developed by George Kelly in 1955. It is often considered to be a theory far ahead of its time, which provided a revolutionary alternative to the behaviorist and psychodynamic approaches prevailing at the time (Chiari & Nuzzo, 2003; Chiari & Nuzzo, 2010; Fransella & Neimeyer, 2003). Kelly (1955) objected to the notion that the individual is "a passive respondent to environmental events" (Fransella & Neimeyer, 2003, p. 25). Also, he

rejected the idea that one's behavior is primarily governed by "internal unconscious forces" (Fransella & Neimeyer, 2003, p. 25) or her past (Kelly, 1955; Kelly, 2003). Instead, he proposed that individuals actively and continually seek to make meaning of life based on their unique interpretations of their experiences. In other words, it is not the events that shape the individual, but rather, the way the individual interprets them. Thus, personal construct theory suggests that individuals are both accountable for and free to create their own future (Kelly, 1955; Kelly, 2003).

Kelly (1955) coined the term "constructive alternativism" to describe the underlying philosophical position of his theory. This philosophical perspective emphasizes the uniqueness of each individual and the possibility of change in how she/he views the world. It holds that although external realities exist, each person interprets them differently, creating her/his own unique version of reality. However, since life is a dynamic process, "all of our present interpretations of the universe are subject to revision or replacement" (Kelly, 1955, p. 15). Kelly (1955, 2003) argued that individuals constantly evaluate the efficiency of their interpretations in the face of new experiences, which allows them to make alterations in and evolve their understanding of the world, and thus, he viewed his theory as "a theory of man's personal inquiry-a psychology of human quest" (Kelly, 2003, p. 3). He illustrated this point with the well-known metaphor of "man-the-scientist." According to Kelly (1955), "every man is, in his own particular way, a scientist" (p. 5). He proposed that just like a scientist, in order to predict and control events, every individual develops hypotheses, which she/he tests by experimentation and finally reviews and/or revises if necessary (Bannister 2003a; Kelly, 1955).

In personal construct theory, these hypotheses, which we rely on to navigate through life, are termed constructs. They shape our view of the world, the way we interpret our current experiences and determine what we expect to happen in the future. They are unique to each individual, however, since they are built from past experiences, and different individuals often experience events differently. This inherently unique nature of constructs is the reason why "people respond to the same situation in very different ways" (Bannister, 2003a, p. 34).

Kelly (1955) explained the details of his theory by means of one fundamental postulate and its eleven corollaries.

The fundamental postulate states that "A person's processes are psychologically channelized by the ways in which he anticipates events" (Kelly, 1955, p. 46).

Kelly regarded the individual as in constant motion, constantly changing and evolving. However, he suggested that each person moves in a particular direction for a reason and in a structured manner rather than randomly. The reason, in his view, is the innate human desire to foresee future events, which the individual tries to achieve through the mental representations of life she/he has invented.

In simple terms, this postulate holds that people's thoughts, emotions, perceptions and behavior are dictated by the personal constructs that they have developed, which they continually experiment with throughout their lives in order to predict and control future events.

The eleven corollaries that Kelly (1955) put forward to elaborate the fundamental postulate are as follows:

1) Construction corollary: "A person anticipates events by construing their replications" (p. 50).

This corollary assumes that personal constructs are developed as a result of repeated experience. It is suggested that when an individual perceives recurrent themes in a series of events, this culminates in her/him developing a general idea, that is, a construct which subsumes the perceived similarities between the events in question. However, Kelly (1955) also pointed out that in order to identify how two things are alike, one also has to distinguish how they are different from other things. Thus, a construct indicates not only the perceived similarities between events, but also the perceived differences between events that are perceived to be similar and others.

2) Individuality corollary: "Persons differ from each other in their construction of events" (p. 55).

This corollary emphasizes the role of individual differences in how people interpret events. It suggests that different people interpret the same event in different ways due to individual differences and the different aspects that they focus on as they process the event.

3) Organization corollary: "Each person characteristically evolves, for his convenience in anticipating events, a construction system embracing ordinal relationships between constructs" (p. 56).

Individual constructs do not exist in a vacuum. Instead, they are organized into a hierarchical system, in which some constructs subsume others. The function of the system is to enable the individual to make predictions that are consistent with each other.

Each person's construction system is unique just like the constructs that they hold and indeed it is essentially the differences in the organization of constructs that cause each one of us to have a unique personality. Also, a construction system is dynamic like individual constructs, which allows us to make any necessary changes to it.

4) Dichotomy corollary: "A person's construction system is composed of a finite number of dichotomous constructs" (p. 59).

Reiterating that personal constructs are developed as a result of repeated experience and that identifying how two things are similar requires an understanding of how they are different from others, Kelly (1955) asserted that every construct is dichotomous or bipolar in nature. He explained that a construct indicates how "two things are alike and different from a third" (Kelly, 1955, p. 111) and thus "in its minimum context, a construct is a way in which at least two elements are similar and contrast with a third" (Kelly, 1955, p. 61).

5) Choice corollary: "A person chooses for himself that alternative in a dichotomized construct through which he anticipates the greater possibility for the elaboration of his system" (p. 64).

This corollary places emphasis on Kelly's view that people can create their own destiny and suggests that we focus on the pole of a dichotomous construct which we believe will help us make better predictions. This may result in the extension or constriction of our system.

6) Range Corollary: "A construct is convenient for the anticipation of a finite range of events only" (p. 68).

This corollary simply states that a construct is limited in terms of the number of events that it can be applied to, that is, no individual construct can be comprehensive enough to be applicable to all situations.

7) Experience corollary: "A person's construction system varies as he successively construes the replications of events" (p. 72).

A construct is tested through experience. While the confirmation of predictions through repeated experiences leads to the consolidation of the construct, disconfirmation results in revision or replacement, depending on how much meaning is attached to the construct in question. The greater the discrepancy between what is expected and what actually transpires, the more disconcerting the experience will feel and the individual will adapt the construct in a way that better suits her/his needs. The

changes in individual constructs eventually lead to a change in the construction system, which may result in the consolidation or the disruption of the system (Bannister 2003b; Kelly, 1955; Kelly, 2003).

8) Modulation corollary: "The variation in a person's construction system is limited by the permeability of the constructs within whose ranges of convenience the variants lie" (p. 77).

Kelly (1955) defined the permeability of a construct as "the capacity to embrace new elements" (p. 80). He pointed out that in order to make sense of new and unexpected events, one needs to have a construction system which allows such events to be incorporated into the system. Otherwise, the experience will simply be lost on the individual and thus the experience cycle will fail to achieve its purpose (Kelly, 1955; Kelly, 2003).

9) Fragmentation corollary: "A person may successively employ a variety of construction subsystems which are inferentially incompatible with each other." (p. 83).

Kelly (1955) noted that a construction system is not necessarily "logic-tight" (p. 85). He explained that an individual may possess construction subsystems which are not consistent with each other, which may lead to inconsistent behavior (Kelly, 1955; Kelly, 2003).

10) Commonality corollary: "To the extent that one person employs a construction of experience which is similar to that employed by another, his processes are psychologically similar to those of the other person" (p. 90).

This corollary is like the other face of the individuality corollary. It proposes that although no two people can have exactly the same constructs, much less the same construction system because of the inherently unique ways in which they interpret events, they may develop similar constructs due to the similarities in their interpretations of events and may thus, think, feel and behave in similar ways.

Kelly (1955, 2003) underscored that it is not the similarities between events themselves that lead to similar constructions, but rather the similarities in the ways different individuals construe events, and he pointed out that different people can thus develop similar constructs and display similar behaviors although they might have experienced different life events.

11) Sociality corollary: "To the extent that one person construes the construction processes of another, he may play a role in a social process involving the other person" (p. 95).

Simply put, in order for people to understand each other and have social relationships, they need to look at the world through each other's eyes, that is, each other's construction of events.

Kelly (1955, 2013) contended that people's thoughts, feelings and behaviors are dictated by their personal constructs. However, he also suggested that one's personal constructs are not carved in stone. Kelly (1955) considered life as a dynamic process and each person as "a form of motion" (p. 48). Thus, he argued that people change and evolve all the time in order to deal with the complexities of life and the changes they face as they move through life.

Kelly (1955), however, proposed that there are certain favorable and unfavorable conditions that support or hamper change. The favorable conditions are the use of fresh elements, experimentation, and availability of validating data. The use of fresh elements involves creating a new context which can be used as the basis for the generation of new constructs. Experimentation refers to the actions one carries out in order to try out a new construct, and validating data to experiential evidence that confirms one's predictions. Kelly (1955) suggested that the process of change is likely to be inhibited unless these conditions are sustained, and noted that there are conditions which are particularly detrimental to such a process. According to Kelly (1955), these unfavorable conditions are threat, preoccupation with old material and no laboratory. In simple terms, threat refers to the tension one feels due to new experiences which are incompatible with her/his core constructs, that is, basic constructs which one relies on for the sustenance of her/his identity and existence (Kelly, 1955). Kelly (1955) pointed out that such incompatibilities will often cause the individual to fall back on core constructs and cling on even more tightly to them. As for the second unfavorable condition on his list, Kelly (1955) explained that too much focus on former experiences hinders the development of new constructs as old experiences are normally viewed through the lenses of deep-rooted old constructs, and underscored the necessity of integrating old and new experiences together in order for new constructs to emerge which accommodate both the old and the new experiences appropriately. The last condition listed by Kelly (1955) puts further emphasis on the necessity of experimentation in the formation of new constructs. It contends that it is impossible for a given target construct to emerge and become an established part of the system in the absence of an environment where the individual can try it out.

With respect to how change takes place, Kelly (1955) put forward the notion of tightening and loosening. According to Kelly's definition (1955), when a construct tightens, it results in clearly defined notions and yields "unvarying predictions," whereas when it loosens, it "permits flexible interpretation and yields "varying predictions." (p.p. 483-484). Thus, in order for new constructs to emerge, it is necessary that first the existing constructs loosen up and become flexible enough to permit alternative interpretations. Change will occur when this loosening phase is followed by the tightening of the newly emerging construct.

Kelly (1955) complemented his comprehensive theory with a methodological tool called the Repertory Grid. Just like the other aspects of personal construct theory, it is based on the fundamental postulate which assumes that people think, feel and behave based on their personal interpretations of events. In line with this basic assumption, it aims to understand individuals' subjective realities by uncovering their personal constructs. Thus, the repertory grid technique is not separate from personal construct theory. Rather, it is directly related to the theory and essentially functions as a means of putting it into practice.

Overall, there seem to be three notions that are core to personal construct theory: control, individuality and change. Above all, obviously, according to personal construct theory, our ultimate purpose in life is to control future events, which we try to achieve by making predictions based on our past experiences. However, we all interpret events differently and that is what is meant by individuality. Kelly (1955, 2003) noted that although there may be similarities between the interpretations of different people, no two people can have identical constructions. Also, according to Kelly (1955, 2003), there is no cause or force other than individuals themselves that leads them to their interpretations. Therefore, nobody and nothing other than ourselves could be held responsible for how we construe our experiences. With respect to change, personal construct theory holds that life is a non-linear process full of changes and we often naturally revise and modify our constructs in order to make life seem more predictable and easier to control. Kelly (1955, 2003) also suggested that we can choose to change our constructions of events and thus "nobody needs to paint himself into a corner; nobody needs to be completely hemmed in by circumstances; no one needs to be the victim of his biography" (Kelly, 1955, p. 15).

CHAPTER III

3. METHODOLOGY

3.1. Introduction

This chapter presents the methodology adopted for this study. It describes the research design, the context of the study, the participants, the instruments used for data collection, the data collection procedures, how the data were analyzed, and how issues of trustworthiness and ethical considerations were addressed.

3.2. Research Design

The study adopted a qualitative multiple-case study approach. Qualitative research in general is based on the assumption that "human behavior is context-bound" (Ary, Jacobs & Sorensen, 2010, p. 424). It attaches particular importance to the participants' perspectives and examines human behavior and social phenomena within the unique and natural contexts in which they take place (Ary et al., 2010). Qualitative case studies, in particular, seek to gain an in-depth understanding of a phenomenon by focusing intensely on a single individual or a small of group individuals in real contexts (Ary et al., 2010; Cohen, Manion & Morrison, 2007). Understanding the various factors that cause the individual(s) to behave in the ways they do is of particular concern in case studies, and various data collection methods are used in an attempt to gain as much understanding as possible about why a given individual behaves in a particular way. Thus, the case study approach is particularly useful in situations where a particular phenomenon is context-bound and the researcher aims to explore the complex interactions of different factors that play a role in the occurrence of the phenomenon (Ary et al., 2010; Cohen et al., 2007; Merriam, 1998; Rowley, 2002).

The decision to adopt a qualitative case study approach for this study was made based on the prevailing view in the literature that washback is a complex phenomenon which often occurs as a result of a complex interplay between factors related to test participants and other contextual factors (Alderson & Wall, 1993; Cheng, 2004; Wang, 2010; Watanabe, 2004a). Multiple cases were used based on the understanding that "studying multiple units can provide better illumination" and enhance the trasferibility of the study (Ary et al., 2010, p. 456; Merriam, 1998; Rowley, 2002; Shenton, 2004).

3.3. Context of the Study

The present study was conducted at the preparatory school of a university in Turkey. The university in question is an English-medium university and all students who have been accepted into a four-year degree program are required to certify their proficiency in English or to take the proficiency test administered by the university shortly before the academic year starts. Students who are unsuccessful in the proficiency test have to complete the year- long English program provided at the preparatory school. Those students are placed into different groups, from beginner to pre-intermediate, based on their scores on the placement test, which is administered after the proficiency test.

The main aim of the preparatory school is to equip the students with the language competences that will enable them to follow their courses in their degree programs. For this purpose, it offers an intensive program of 24 to 30 hours of English a week at three levels: Beginner, Elementary and Pre-intermediate. The program consists of 3 courses: Coursebook, Reading and Writing, and Listening and Speaking (and ESP only in the second semester). Each group is assigned 2 non-native speaker teachers of English and 2 native speaker teachers of English. The two non-native teachers teach Coursebook. They teach the same coursebook and follow the same pacing schedule. The native speaker teachers teach Reading and Writing and Listening and Speaking courses. They teach different coursebooks and follow different pacing schedules. In a nutshell, there are 3 different courses taught by 4 different teachers. Each course follows a different coursebook and a different pacing schedule. What these three different courses have in common is that the pacing schedule for each course is strictly based on the coursebook used and designed by the group coordinator (see Table 1).

The assessment system is made up of a complicated network of progress tests, performance assessments and an end-of-year test administered at the very end of the academic year. The progress tests consist of vocabulary quizzes, pop quizzes and monthly tests. The monthly tests are basically extended versions of the pop-quizzes in terms of content, format and the task types they include. However, unlike the quizzes, the monthly tests test students on material covered in class in all of the three courses, and therefore they consist of 4 different sections: Coursebook, Reading, Writing and Listening. Neither the quizzes nor the monthly tests assess speaking skills. The end-of-year examination, that is the final exam, is basically a more comprehensive version of

the monthly tests, which also includes the testing of speaking skills. In addition to these tests, students are also assessed and graded on their presentations in Listening and Speaking classes and on their essays in Reading and Writing classes.

In its simplest form, the system is made up of 87 tests (53 in the first semester and 34 in the second semester). When the performance assessments administered in Listening and Speaking and Reading and Writing classes are added, the number of exams goes up to 110. Interestingly enough, there were 160 working days in the given academic year, which means an exam was administered at the school every 1.45 days. Given these startling numbers, along with the fact that all of these examinations were designed mainly by the course and group coordinators under the supervision of the administration, it is easy to see the power that the system assigns to exams (see Table 2).

As in most other academic settings, the final exam has greater weight (it accounts for 60% of the final grade) on the final grade than all the other exams and because of its power to determine whether or not a student can begin her/his degree program, it may be considered as a high-stakes test. However, this test is typically a comprehensive replica of the monthly tests, which are a comprehensive and extended version of the quizzes. Since the road to the final exam is paved with the preceding 80 quizzes and 6 monthly tests, it is those exams that teachers and students are faced with the most frequently and talk about the most. Indeed, these tests are administered so frequently and are such a great part of the teachers' and students' lives that they often receive much more attention than does the end-of year-examination until towards the end of the academic year. For this reason, this present study focused on the influence of the quizzes and monthly tests, rather than the final examination or the performance assessments.

Table 1.

Courses and levels taught at the institution

LEVELS	COURSES			
	COURSEBOOK	LISTENING & SPEAKING	READING & WRITING	ESP
Pre-Intermediate	•1st & 2nd semesters •14 hours a week •by Turkish teachers of English	•1st & 2nd semesters •5 hours a week •by native-speaker teachers of English	•1st & 2nd semesters •5 hours a week •by native-speaker teachers of English	•2nd semester •2 hours a week •by Turkish teachers of English
Elementary	•1st & 2nd semesters •16 hours a week •by Turkish teachers of English	•1st & 2nd semesters •5 hours a week •by native-speaker teachers of English	•1st & 2nd semesters •5 hours a week •by native-speaker teachers of English	•2nd semester •2 hours a week •by Turkish teachers of English
Beginner	•1st & 2nd semesters •18 hours a week • by Turkish teachers of English	•1st & 2nd semesters •5 hours a week •by native-speaker teachers of English	•1st & 2nd semesters •5 hours a week •by native-speaker teachers of English	•2nd semester •2 hours a week •by Turkish •teachers of English

Table 2.

Number of exams

	FIRST SEMESTER	SECOND SEMESTER
COURSEBOOK	Vocabulary quizzes: 14 Pop-quizzes: 8 TOTAL: 22	Vocabulary quizzes: 7 Pop-quizzes: 6 TOTAL: 13
LISTENING & SPEAKING	Vocabulary quizzes: 10 Pop-quizzes: 2 Presentations: 10 TOTAL: 22	Vocabulary quizzes: 10 Pop- quizzes: 2 Presentations: 10 TOTAL: 22
READING & WRITING	Vocabulary quizzes: 14 Pop-quizzes: 2 TOTAL: 16	Pop quizzes: 3 Essays: 3 TOTAL: 6
ESP	-	Quizzes: 2 TOTAL: 2
MONTHLY TESTS	3	3
FINAL EXAM	-	1
TOTAL	63	47

3.4. Participants

The participants in this study consisted of 3 English teachers working at the preparatory school of the university. Two of them were non-native speakers of English and one was a native speaker of English. None of them was a coordinator.

The purposive sampling technique was used to gather participants who would potentially have differing beliefs and thus to "provide maximum insight" (Ary et al., 2010). Based on the understanding that experience plays a major role in the formation

of beliefs and perceptions, three major parameters were set for the selection of participants: 1) degree of teaching experience 2) length of employment at the institution and 3) educational background. However, taking into account that beliefs become established over time as a result of repeated exposure to similar events and assuming that it would take time to fully integrate into a new system, two other criteria were set by the researcher: (1) having at least 3 years of teaching experience and (2) having completed at least one year of employment at the institution.

In order to find participants that fit all the criteria, initially, the researcher reviewed the resumés of the teachers working at the preparatory school published on the university's website. Then, she talked to 15 teachers who fit the criteria to verify the accuracy of the information provided on the website. From among those 15 teachers, three teachers who were willing to participate took part in the study.

 For confidentiality purposes, each participant was assigned a pseudonym (Emma, Sarah and Jason).

3.5. Instruments

3.5.1. The Repertory Grid Technique

The repertory grid technique (rep-grid) was used to identify the participants' beliefs about ideal teaching practices (see Appendix 2 for a sample repertory grid). The rep- grid technique is an assessment technique originally developed by Kelly (1955) for the elicitation of personal constructs. It mainly aims to uncover people's personal beliefs and perceptions about a given topic based on their unique interpretations of elements that are representative of the topic concerned. As teachers' beliefs are conceptualized in this study within the framework of personal construct theory, the participant teachers' beliefs about ideal teaching practices were elicited via the rep grid technique.

Basically, a repertory grid consists of three main components: elements, constructs and "a linking mechanism" (Smith, 1980, p. 4). Elements refer to people, things or events that are representative of a given topic. According to construct theory, "in its minimum context, a construct is a way in which at least two elements are similar and contrast with a third" (Kelly, 1955, p. 6). Thus, a set of three related elements is required for the elicitation of each construct. The similarity between the two elements is

called the emergent or similarity pole and the way the two elements contrast with another element is called the implicit or the contrast pole (Kelly, 1955).

"Linking mechanism" refers to the scaling system used in a rep-grid to show how the constructs and elements are related. There are three systems used for this purpose: dichotomizing, rating scales and rankings (Smith, 1980). In this study, a five-point rating scale was used.

3.5.2. Classroom Observations

A total of 12 hours of direct non-participant classroom observations were conducted in order to examine whether the participants' identified beliefs were "reflected in their behavior" (Alderson & Wall, 1993, p. 127). Each participant was observed for 4 teaching hours. Each observation was audio-recorded, and detailed field notes were taken during the observations.

3.5.3. Post-observation Interviews

An in-depth semi-structured interview was conducted with each participant following the classroom observations. The questions were designed based on an indirect approach in order to be able to elicit honest responses from the participants (Cohen et al., 2007). Each interview originally consisted of 5 questions which were intended (1) to complement the rep grid data by gathering further information about the teachers' beliefs about ideal teaching practices, (2) to gather information about the teachers' perceptions of the tests, and (3) to help gain an in-depth understanding of why exactly the teachers did what they did in the classroom and to find out whether and how it was related to the exams. Each interview lasted 48-64 minutes and was audio-recorded (see Appendix 3).

3.6. Procedure

Data were collected in three interrelated stages. As each stage was designed to complement another stage, it was expected that data collected in one stage would not only feed into the data collected in another stage but would also enlighten the researcher as to the kinds of details she should pay attention to in the next stage. Taking this into account, the researcher analyzed each set of data immediately after it was collected

rather than waiting until all the data had been gathered (Ary et al., 2010). Thus, in this study, data analysis was an important part of the data collection process.

The first stage of the study aimed to identify the teachers' beliefs about ideal teaching practices via the rep grid technique. The constructs were elicited from the participants themselves using the traditional triadic elicitation technique (Bell, 2003). First of all, in order to help the participants understand the procedure, each participant was given a blank rep grid, and the researcher demonstrated how to fill out the grids through a sample. She asked the participants to think of three teachers they considered effective, three they considered average and three they considered ineffective in terms of their teaching practices, then to code (E1, E2, E3, A1, A2, A3, I1, I2, I3) and write them at the top of the matrix. For the next stage, they were asked to think how any two of the teachers were similar to each other and different from a third, and to write the similarities on the left side of the matrix, that is, the emergent pole, and the differences on the right side of the matrix indicating the contrast pole. Finally, the participants were asked to rate each element on a 5-point scale based on how close they thought each element was to the emergent or the contrast pole. The participants completed the grids at their convenience, and immediately after all of them returned their grids, the researcher began analyzing the data. When the analyses were completed, the findings were triangulated through member checking.

In the second stage, a total of 12 hours of direct non-participant classroom observations were conducted over a period of 4 weeks in order to examine whether the teachers' classroom practices were consistent with their identified beliefs. Each teacher was observed on two different occasions. On each occasion, the teacher was observed for two successive teaching hours (see Table 3). Each observation took place at a pre-arranged date and time. During the observations, the researcher took detailed field notes. Guided by the findings from the rep grid data, she paid particular attention to whether the teachers' teaching methodologies reflected their identified beliefs and made notes of any inconsistencies that occurred. Also, each observation was audio-recorded.

The classroom observations were followed by in-depth semi-structured interviews. The participants were interviewed separately and at a date and time that was convenient to them. The Turkish teachers were interviewed in Turkish and the teacher who was a native speaker of English was interviewed in English. The interviews that were conducted in Turkish were later translated into English. Each interview lasted 48-64 minutes and was audio-recorded. The overall purpose of the interviews was to build

a holistic understanding of the teachers' perspectives. Thus, the researcher made a deliberate effort to make the participants feel at ease. She did her utmost to create a comfortable and friendly atmosphere so that the participants would feel comfortable sharing their genuine opinions.

Table 3.

Participants observed and the number of observations conducted over 4 weeks

	Week 1	Week 2	Week 3	Week 4	Total
Emma	2 lessons	2 lessons			4
Sarah	2 lessons		2 lessons		4
Jason			2 lessons	2 lessons	4
					12

3.7. Data Analysis

As stated before, in this study data collection and data analysis were closely related. Initially, each set of data was analyzed separately for each case. Subsequently, the findings from different sets of data collected for one case were compared with each other to gain a holistic understanding of the case. Finally, all the cases were cross analyzed to build a holistic understanding of all the findings.

3.7.1. Cluster Analysis

Cluster analysis was used for the analysis of rep grid data. Cluster analysis uses the FOCUS program, which yields a FOCUsed grid showing the degree of similarity between constructs or elements (Shaw & Thomas, 1978). Initially, each participant's rep grid data were entered into WebGrid 5 for cluster analysis. Then, the participants were asked to check whether the results reflected their pedagogical beliefs. All the participants confirmed the accuracy of the results. Thus, no further analysis was required.

3.7.2. Content Analysis

Data collected through classroom observations and post-observation interviews were analyzed using content analysis. In line with the overall design and purposes of the study, a qualitative approach was adopted, and the categories were generated inductively from the data. "Qualitative content analysis usually uses individual themes as the unit for analysis, rather than the physical linguistic units," thus, codes were created based on emergent themes regardless of the size of the text (Zhang & Wildemuth, 2009, p. 310). While coding the data, both manifest and latent content were taken into account, and particular attention was paid to the context while coding each set of classroom observation data (Zhang & Wildemuth, 2009).

The coding process began immediately after the collection of each set of data. The classroom observation data and post-interview data were analyzed separately for each case. For classroom observations, audio-recordings were fully transcribed, and the transcripts of the audio-recordings and the field notes were analyzed as a single unit of text. Once the coding was completed, the findings were matched against those from the rep-grid data to identify whether there were any inconsistencies between the teachers' identified pedagogical beliefs and their actual classroom practices, which also fed into the post-observation interviews. The audio-recordings of the post-observation interviews were also fully transcribed and analyzed in the same way as the classroom observation data.

3.8. Trustworthiness

Trustworthiness is the qualitative counterpart of reliability and validity in quantitative studies, and it consists of credibility, transferability, dependability and confirmability (Guba, 1981; Shenton, 2004). In this study these issues of trustworthiness were addressed mainly through the triangulation of different data sources, member-checking, the use of multiple cases and thick and detailed descriptions. In order to address issues of credibility and confirmability, three different types of data were collected for each case and matched against each other to check for consistency and to "reduce the effect of investigator bias." (Shenton, 2004, p. 72). Also, the accuracy of the findings was checked through member-checking at different stages of the study to increase the credibility of the findings. The issues of transferability and

dependability were addressed through the use of multiple cases, and thick and detailed descriptions of the cases, contextual factors and the methodological processes (Guba, 1981; Patton, 1999; Shenton, 2004; Zhang & Wildemuth, 2009).

3.9. Ethical Considerations

Participation in this study was on a voluntary basis. Prior to the study, all of the participants were informed about the purpose, the instruments to be used and the study procedures verbally. They were also informed that their real identities would not be revealed. Then, they were asked to sign an informed consent form (Appendix 1).

CHAPTER IV

4. FINDINGS

4.1. Introduction

This chapter presents the findings that emerged from the data analysis. The findings are presented in three sections separately for the three cases. Each section begins with a brief description of the case participant, continues with the presentation of the findings separately for each set of data and ends with a summary of the relevant findings.

4.2. Case 1- Emma

Emma is a Turkish-speaker teacher of English with a total of 21 years of teaching experience. She holds a Bachelor's and a Master's degree in ELT. At the time of the observations, she had been working at the institution for nine years and six months, and was in her tenth academic year. She was teaching Coursebook to two different classes at the elementary level and ESP to law students, her total number of teaching hours amounting to 18 hours per week. She was not a coordinator. The observations took place in one of her Coursebook classes.

Table 4 presents the general characteristics of Emma along with the courses and the levels she was teaching at the time of the observations.

Table 4.

Emma-General characteristics and the courses/levels she was teaching at the time of the observations

Native language	Turkish
Formal Training in ELT	BA and MA in ELT
Total years of teaching experience	21
Length of employment at the institution	In her 10th academic year
Courses/Levels she was teaching	Coursebook/Elementary & ESP
Courses/Levels observed	Coursebook/ Elementary
Total number of teaching hours per week	18
Coordinator	No

4.2.1. Results from the Cluster Analysis of the Rep Grid Data

Emma generated 15 constructs about ideal teaching methods and the cluster analysis of her constructs produced three main clusters (see Figure 1). The largest one of these clusters consists of eight constructs, among which "use of icebreakers" versus "no use of icebreakers" and "a good sense of humor" versus "no sense of humor" as well as "has enough knowledge of subject" versus "doesn't have enough knowledge of subject" and "speaks English in class" versus "doesn't speak English in class" are matched at over 90%, and the construct "integrates the four skills" versus "doesn't integrate the four skills" is linked to that pair at about 90%. "Uses body language/ gestures" versus "doesn't use body language/ gestures" appears as a rather isolated construct in the cluster whereas the two other constructs seem to be linked to the first pair at 90% and over 80% although they do not form a pair.

Figure 1. *Cluster analysis of Emma's constructs about ideal teaching methods*

The second cluster consists of two tight pairs. The first one is "well-prepared for class" versus "not well-prepared for class" and "uses the communicative approach" versus "uses traditional methods," which are matched at over 85%. The second pair in this cluster is, "uses pair work/ group work activities" versus "doesn't use pair work/ group work activities" and "monitors students during the lesson" versus "doesn't monitor students during the lesson," which show a similarity score of over 90%.

The last two constructs on the grid form a distinct cluster, with a similarity score of 80%.

The relationships between the constructs in the first cluster suggest that Emma believes a positive classroom climate is the foundation for positive classroom interactions, which mainly requires the teacher to reduce tension and stress in the classroom through a good sense of humor. Also, she seems to believe that the teacher should interact with the students in various ways, but ideally in the target language, which requires a good command of the language on the part of the teacher. Taking into account that the construct "integrates the four skills" is associated with the pair that consists of "speaks English in class" and "has enough knowledge of subject," it could be suggested that she believes teacher talk in the target language is a part of skills integration.

The second cluster clearly shows that she believes using communicative language teaching methods entails thorough pre-class preparation on the part of the teacher. The relationship between "uses pair work/ group work activities" and "monitors students during the lesson" suggests that Emma believes in-class pair and group activities should be carefully monitored by the teacher.

The last cluster that emerged from the analysis indicates that she believes rules should be presented inductively in written form by means of effective use of the board.

The isolated construct, "gives students lots of exercises" is self-explanatory. It simply shows that Emma believes a teacher should ideally give a lot of exercises. However, it is not associated with any of the other constructs in her grid.

Overall, these results suggest that Emma has knowledge about and interest in CLT. However, the fact that certain methods and techniques associated with CLT such as the integration of four skills and inductive teaching did not appear in the same cluster as the constructs "uses the communicative approach" and "uses pair work/ group work activities" suggests that she has not established close links between these constructs and CLT.

4.2.2. Observation Results

4.2.2.1. Overview of Emma's Lessons

As stated before, Emma was observed in one of her Coursebook classes for four hours on two different occasions and on each occasion, she was observed for two successive teaching hours as were all the other participants. At the most general level,

each two-session block could be described as being characterized by three stages: a pre-lesson warm-up stage, the main teaching and learning stage and a warm-down stage. The pre-lesson warm-up stage consisted of a brief chat with the students. The second stage was characterized by an undivided focus on the coursebook, and the warm-down stage consisted of a song lyrics gap-fill activity. The language of instruction was English in all of the stages in all of the lessons except for occasional one or two-word translations of new or unknown words.

Apart from the warm-down activities, the main teaching material was the coursebook. Following the pre-lesson warm-up stage, Emma briefly reminded the students about the page at which they had left off in the previous lesson and focused her attention on the coursebook. This stage was strictly based on the content of the coursebook. Typically, she presented the target language functions inductively in a relevant context, which she did either through a teacher-produced short conversation or the texts provided in the coursebook. She also held whole-class discussions based on the discussion questions provided in the book. At the beginning of such discussions, she had the students brainstorm ideas as guided by the coursebook and focused on their own culture and experiences throughout the discussions. For vocabulary teaching, she employed different techniques such as English definitions, Turkish translations and example sentences, and she seemed particularly careful about the specification of the parts of speech. She systematically wrote every explicitly taught item on the board such as the language functions and the target vocabulary. Additionally, she made use of the board during the brainstorming activities. For the speaking and writing tasks in the book, she had the students work in pairs. While the students were working on the tasks, she monitored them carefully, providing immediate corrective feedback for grammar mistakes and answering each individual student's questions in detail.

The types of feedback she gave the students mainly consisted of oral evaluative and corrective feedback. Indeed, evaluative feedback seemed to be an integral part of the way she interacted with the students. She gave profuse positive evaluative feedback to students throughout the lessons, regardless of their responses or their mistakes. As for corrective feedback, she corrected errors both explicitly and implicitly during whole-class discussions whereas she typically provided explicit error correction while monitoring pair work activities. A particularly noteworthy characteristic of her teaching was her focus on accuracy throughout the lessons. During all the speaking activities, she tended to correct every incorrect utterance produced by the students, giving the

impression that she was more interested in helping students produce correct sentences than she was in fluency or authentic interaction. Another feature of her teaching that is definitely worth mentioning was the remarkably polite language she used while interacting with the students, which seemed to be a part of her efforts to make the students feel at ease in class.

4.2.2.2. Results from the Content Analysis of the Classroom Observation Data

The content analysis of the field notes and transcripts of the audio-recordings yielded four major categories, which showed that Emma's teaching was characterized by a combination of CLT and traditional methods (see Table 5). Clearly, she made use of methods and techniques that are typically associated with CLT such as inductive teaching, teaching language in context, pair work, use of the target language, integration of language skills, use of authentic materials and relating new information to students' own lives. Also, her attentiveness to individual students was suggestive of the importance she attached to the needs of individual students, which is in line with the underlying philosophy of CLT. On the other hand, another key component of her teaching was extensive focus on accuracy at the expense of fluency, which is a characteristic of traditional approaches (Hall, 2011; Richards, 2006; Richards & Rogers, 2001).

When compared with the findings from the rep grid data, it was revealed that her use of CLT methods and the positive evaluative feedback she gave the students, the extra-syllabus activities she incorporated into the lessons and her frequent use of the board for instructional clarity were consistent with her identified beliefs about ideal teaching methods. However, considering her belief that CLT is the ideal approach to teaching, her focus on accuracy at the expense of fluency and authentic communication emerged as a major inconsistency between her beliefs and her actual classroom practices (Hall, 2011; Richards, 2006; Richards & Rogers, 2001).

This major divergence between one of Emma's beliefs and her actual classroom practices was noted to be discussed with the participant in the post-observation interview.

Table 5.

Emma-Results from the content analysis of the classroom observation data

Code	Example/Quote	Category
Contextual teaching of language functions	"Imagine that you want to invite your friend somewhere. You can say 'Are you free today?' And the other person says 'I'm busy.' Or, another person says 'Do you want to go for coffee?' and the other person says 'No, sorry. I have got too much to do.'"	
Inductive teaching of language functions	"Before we invite somebody somewhere, we ask a question: 'Are you free today?' Why do we ask this question?" (Students try to guess)	
Integrated skills	Listening & speaking tasks in the book and class discussions Reading and writing exercises in the book & class discussions	
Stimulation of students' background knowledge	"Now, tell me, what methods can you use to improve your listening?" In İstanbul, do you think neighbors chat with each other a lot?"	
Making learning relevant	"Do you live in a dormitory or in a flat?" "Have you ever fallen in love?"	**(1) Combination of CLT and traditional Approach**
Student collaboration	Pair work for a speaking task Pair work for a writing task	
Attentiveness to individual students	Walks around the class as students work on a speaking/writing task and checks each pair's paper giving immediate corrective feedback and answering their questions in detail.	
Use of authentic material	Song lyrics activity	
Focus on accuracy rather than communication	Teacher: "What do you do with your neighbors? Student: "We are having barbecue". Teacher: "You *have barbecues"* (moves on to another student) Teacher: "What can you do to improve your listening?" Student: "We can friends with foreigners". Teacher: "We can ??? friends. There's a word there. You learned it last week. We can?? *make* friends" (moves on to another student).	

Table 5 (continued)

English definitions	"Suburb is an area of houses in outer parts of a city or town."	
Turkish translations of new words during class discussions	"In Turkish, 'mercy' means 'merhamet.'"	
Target vocabulary on the board	Writes all the target vocabulary on the board	**(2) Explicit Vocabulary Instruction**
Oral/ Written specification of parts of speech	Specifies whether each word is a verb, adjective, noun etc "Complain is a verb. Complaint is the noun form."	
Teacher modelling	"For example, the distance between İstanbul and Ankara is about 450 km."	
Writes vocabulary on the board	Writes all the explicitly taught vocabulary on the board	
Writes teacher explanations on the board	Writes teacher explanations about language functions on the board	**(3) Use of the board to support clarity of instruction**
Writes students' ideas on the board	Writes most students' ideas on the board while brainstorming ideas	
Positive evaluative feedback	"Very good"(regardless of mistakes) "Excellent"(regardless of mistakes)	**(4) Positive classroom environment**
Very polite language	"OK, ladies and gentlemen, could you look at page 87, please?"	

4.2.3. Post-observation Interview Results

The interview with Emma lasted 48 minutes and took place in her office at the institution. In addition to the original five main interview questions, she was asked follow-up and probe questions depending on the direction of the conversation and the topics that emerged during the interview (see Appendix 3). She was also asked questions aimed at probing into the possible reasons for the inconsistency between one of her beliefs and her actual classroom practices. However, her belief regarding giving lots of exercises emerged as a latent theme on its own in response to the fifth question and was interpreted within the context of that question. The content analysis of the interview transcript yielded eight main categories (see Table 6).

The results suggest that Emma believes all language skills are interrelated and that because of this interrelationship, an integrated-skills approach is necessary to effect effective learning. However, she also seemed to believe that vocabulary knowledge and listening skills play a particularly important role in understanding language input.

When she was asked what aspects of the language she focused on the most in her daily practice, she listed grammar, vocabulary and listening as the main areas of focus, as is evident from the following excerpt:

Here, we focus mostly on grammar, but I personally try to focus on vocabulary and listening in my daily practice. I also focus on listening activities. Skills are important.

The same excerpt also seems to have a latent meaning which admits of the interpretation that focus on grammar is standard practice due to the system and that she focuses on vocabulary and listening in spite of the system.

In terms of her teaching methods and techniques, a recurrent theme that emerged from her response to the same question was the similarity between the approach adopted by the coursebook and her preferred methods of teaching. Interestingly, after each method and technique she listed, she referred to the exercises, activities or the teaching style of the coursebook and mentioned how similar they were to how she taught in class, which gave the impression that she considered the similarities between the methods she employed and the style of the coursebook as a justification of the methods she claimed to use.

When she was asked if there was a particular reason for her intense focus accuracy in general and even during speaking activities, she explained:

I do it just out of habit. It's an occupational habit. So that they can make correct sentences.

Because they were going to do a speaking activity, I wanted them to be able to speak making correct sentences.

Upon further probing, she directly stated that it was not related to the exams.

These excerpts suggest that Emma's constant correction of errors was probably an ingrained habit, which probably resulted primarily from her personal beliefs about the importance of accuracy.

Her responses also showed that she believed the exams conducted at the institution had two main purposes: to assess grammar, language functions, vocabulary and other skills and to assess teacher performance. She repeatedly stated that exams were a "gauge of teacher performance." She also seemed to be remarkably sensitive about student test scores:

> *I feel on top of the world when my students get high scores on the exams. I feel as if I have really been able to teach something.*
> *If a student gets a bad score on an exam, I feel as if I haven't been able to teach anything.*

With respect to the washback effects of the exams, she reported that there were too many exams, explaining that having too many exams demoralized the students and caused them to focus only on getting a passing grade rather than encouraging them to practice what they had learned. She further explained that having too many exams resulted in superficial and temporary learning due to limited practice.

It also appeared that the exams exerted negative washback both on teaching content and her teaching methodology. Emma stated that she could not always do extra activities because she had to cover the book first. Her responses suggested that she had to limit certain activities such as pair and group work and eliminate some altogether such as games, vocabulary activities, gap-fill exercises for vocabulary, sentence completion and matching activities due to time constraints. She also complained that she was not able to provide the students with as many follow-up exercises as she would like to because of time constraints.

Obviously, according to Emma, the cause of time constraints was the intense pacing schedule:

> *The pacing schedule is so intense. It causes time constraints.*

When she was asked what would happen if she fell behind the schedule, it was revealed that she had major concerns regarding accountability, which seemed to be the underlying reason for her concerns over time. The following excerpt is a succinct summary of the relationship there seems to be between the pacing schedule, time constraints, Emma's sense of accountability to the administration and potential washback from the exams on her teaching:

> *I have to catch up with the pacing schedule because we have exams. If I fell behind the schedule, my students wouldn't be able to do the questions on the exam and they would complain to the administration about me.*

Table 6.

Emma- Results from the content analysis of the interview transcript

Code	Example/Quote	Category
Interrelated skills	"All skills are interrelated."	
Incorporation of all skills	"I think. In order to effect effective learning, all skills should be incorporated."	
Contextual teaching of grammar	"I generally present each grammar item through a listening activity, a reading activity or some sort of relevant context. Then we discuss the grammar item in context."	**(1) CLT**
Inductive teaching of grammar	"I ask them to think about the rules. To look at the text and find the rule. Sometimes I write an example sentence and elicit the rule from them and write the rule on the board."	
Focus on grammar, Focus on vocabulary Focus on listening	"Here, we focus mostly on grammar, but in my daily practice, I personally also focus on vocabulary and listening."	**(2) Main areas of teaching focus**
Similarity between how she teaches and how the book teaches	"This is also how our book typically teaches it."	**(3) Coursebook as the justification of her methods**
Similarity between the exercises she provides and those provided in the book	"We also have such activities in the book."	

Table 6 (continued)

Code	Example/Quote	Category
Occupational habit	"I do it just out of habit. It's an occupational habit. So that they can make correct sentences."	**(4) Ingrained focus on accuracy**
Focus on correct sentences	"Because they were going to do a speaking activity, I wanted them to be able to speak making correct sentences."	
Testing grammar, Testing language functions Testing vocabulary Testing L & R	"The quizzes test grammar, language functions and production and vocabulary knowledge. The monthly exams test grammar productivity, vocabulary, listening and reading."	
Assessing teacher performance	"Exams also assess teachers. We understand how well we have taught." "If a student gets a bad score on an exam, I feel as if I haven't been able to teach anything." "Student test scores are also a gauge of teacher performance." "Exams are also a test for us."	**(5) Functions of exams**
Temporary learning	"They (the exams) lead to temporary learning. The exams don't support long-term retention of information."	
Lack of student practice	"They (the exams) don't encourage students to learn and practice what they have learned."	**(6) Washback on students and student learning**
Focus on exam scores	"The exams cause the students to focus only on getting a passing grade."	
Demoralization of students	"Having too many exams demoralizes the students and hinders them from internalizing what they have learned."	

Table 6 (continued)

Focus on exam content	"I can't always give enough follow-up exercises. Because we have to finish the book first. Otherwise, I'll fall behind the pacing schedule."	
Elimination of games Elimination of vocabulary gap-fill exercises Elimination of sentence completion Elimination of matching activities Reduction of pair work Reduction of group work activities	"I'd incorporate games, vocabulary activities, vocabulary exercises-gap-fill exercises, sentence completion and matching activities, more pair and group work. I used to do these in the first semester, but I can't this semester because I have to catch up with the pacing schedule."	**(7) Washback on teaching**
Too many exams	"The system here is too much exam-oriented. There are too many exams." "There should be fewer exams. Having too many exams demoralizes the students and hinders them from internalizing what they have learned."	
Time constraints due to intense pacing schedule	"The pacing schedule is so intense. It causes time constraints."	**(8) Causes of washback**
Accountability to the administration	"I *have to* catch up with the pacing schedule because we have exams. If I fall behind the schedule, my students wouldn't be able to do the questions on the exam and they would complain to the administration about me."	

4.2.4. Summary of Relevant Findings for Emma

A comparison of the cluster analysis results and observation results revealed both consistencies and a major inconsistency between Emma's identified beliefs and her actual classroom practices. Her apparent efforts to create a positive classroom environment, her use of English throughout the lessons, the generous amounts of positive evaluative feedback she gave the students, the incorporation of extra-syllabus activities into the lessons, the integration of the four skills, the inductive presentation of the rules by means of the board, the pair work activities she used and her careful monitoring of the students were all reflective of her previously identified beliefs.

However, her extensive focus on grammatical accuracy at the expense of fluency and authentic interaction during the lessons emerged as a major divergence between her belief regarding the necessity of the adoption of a communicative approach to language teaching and her actual classroom practices.

The interview results indicated that the main cause of this inconsistency was Emma's ingrained focus on accuracy rather than the exams, which might have its roots in a belief about the significance of accuracy. With regard to the exams, the interview results showed that Emma believed the exams conducted at the institution served two main purposes: to assess grammar, language functions, vocabulary and other skills and to assess teacher performance. Overall, she seemed to be content with the level of alignment between the tests and what she taught in class. With respect to washback effects, it was found that Emma believed the exams exerted negative washback on both learning and teaching. In terms of the washback effects of the exams on her teaching in particular, it appeared that she had to limit or eliminate some activities and the number of certain exercises as a consequence of time pressures. Evidently, the intense pacing schedule caused time pressures; however, it appeared that her strong sense of accountability to the administration was the root cause of her concerns over time.

4.3. Case 2- Sarah

Sarah is a native-speaker teacher of English with a total of 4 years of teaching experience. At the time of the observations, she had been working at the institution for one year and six months and was in her second academic year. She was teaching Listening and Speaking to four different classes at the beginner, elementary and pre-intermediate levels, her total number of teaching hours amounting to 20 hours per week. She was not a coordinator. The observations took place in one of her beginner classes.

Table 7 presents the general characteristics of Sarah along with the courses and the levels she was teaching at the time of the observations.

Table 7.

Sarah- General characteristics and the courses/levels she was teaching at the time of the observations

Native language	English
Formal Training in ELT	None
Total years of teaching experience	4
Length of employment at the institution	In her 2nd academic year
Courses/Levels she was teaching	Listening & Speaking/ Beginner, Elementary, Pre-Intermediate
Courses/Levels observed	Listening & Speaking/ Beginner
Total number of teaching hours per week	20
Coordinator	No

4.3.1. Results from the Cluster Analysis of the Rep Grid Data

Sarah generated 15 constructs, and the cluster analysis produced one isolated construct and 2 clusters, each consisting of 7 constructs (see Figure 2). Her FOCUSed grid shows that in the first cluster, 6 of the constructs form 3 pairs. Among these pairs, "organized" versus "disorganized" and "relates new concepts to previous ones" versus "doesn't relate new concepts to previous ones" are matched at about 85%. Similarly, "use of humor" versus "threatening and unwelcoming environment" and "learning by doing" versus "memorize and regurgitate" have a similarity score of about 85%. The third pair in the cluster consists of the constructs "encourages group activities" versus "discourages group activities, and "creative and adaptive" versus "endless worksheets/ tests/ homework," which are matched at about 90%. The construct "encourages questions" versus "doesn't encourage questions" stands alone; however, it is linked to the cluster at over 80% match level.

In the second cluster, "motivational" versus "not motivational" and "good assessment and prompt feedback" versus "no feedback in class" form a pair with a similarity score of about 90%, to which "recognizes that students are individuals" versus "mass production standards" and "explanations at appropriate level" versus "explanations not at appropriate level" are linked at about 85%. There appear to be three other constructs linked to the cluster at over 80%; however none of them have formed a pair with each other.
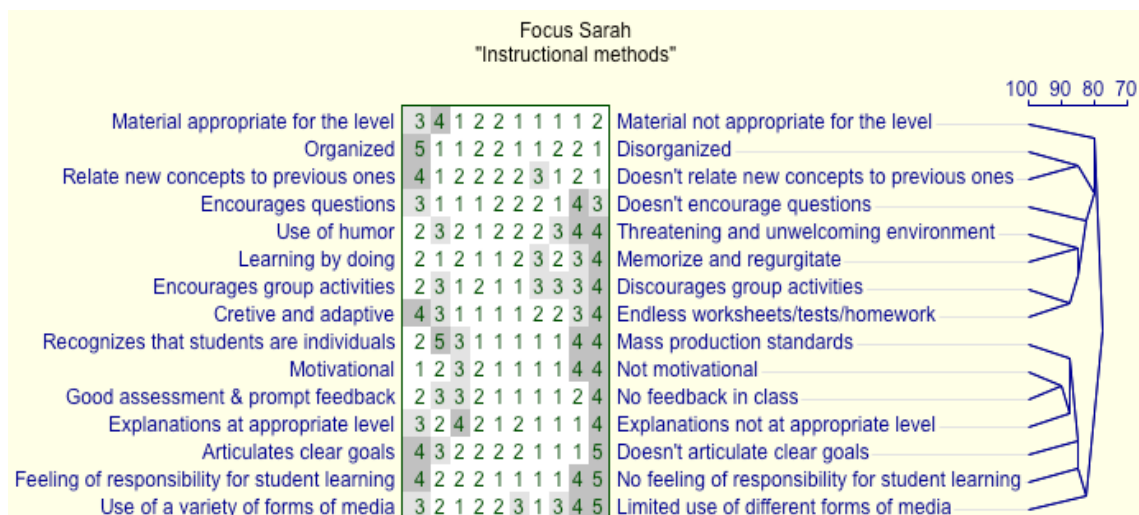
Figure 2. *Cluster analysis of Sarah's constructs about ideal teaching methods*

The isolated construct shows that Sarah thinks it is important to use level-appropriate materials while teaching.

Taken overall, the first cluster in her grid suggests that Sarah believes a teacher should ideally hold a learner-centered perspective and use methods and techniques such as relating new concepts to previous ones, creating opportunities for students to learn by doing, and encouraging questions and group activities. The connections between the constructs show that she believes a teacher needs to be organized to be able to relate new concepts to previous ones. Also, it seems that according to Sarah, use of humor is necessary to support learning by doing; and the teacher needs to be creative and adaptive to be able to run group activities.

The second cluster includes some constructs often associated with formative assessment practices such as "good assessment and prompt feedback" and "articulates clear goals." Also, "recognizes that students are individuals" and "feeling of responsibility for student learning" could be considered as redolent of formative assessment. In the second cluster, the close relationship between "motivational" and "good assessment and prompt feedback" indicates that Sarah believes providing good assessment and prompt feedback will motivate students. Also, "recognizes that students are individuals" and "explanations at appropriate level" are somewhat loosely linked to this pair, based on which it could be speculated that she would to expect feedback to be student-specific and level appropriate in order to be motivational. Nonetheless, the fact that "articulates clear goals" and "feeling of responsibility for student learning" are isolated constructs within the cluster suggests that she has not developed a system that

links them together. Regardless, taken at face value, this cluster suggests that Sarah would expect an ideal teacher to provide good assessment and prompt feedback to be motivational, which is related to recognizing students as individuals and level-appropriate explanations. She also believes articulating clear goals, having a feeling of responsibility for student learning and using different forms of media are important characteristics of effective teaching.

4.3.2. Observation Results

4.3.2.1. Overview of Sarah's Lessons

In all of the lessons observed, teaching content was strictly based on the content of the coursebook, and the learning and teaching activities were primarily guided by the activities and exercises provided in the book. The language of instruction was English from beginning to end in each lesson.

The teaching and learning activities mainly consisted of the articulation of the learning objectives, teacher explanations, whole-class discussions, completion of the exercises in the coursebook and pair and group discussions. Sarah typically started each lesson by briefly reminding the students about the previous lesson. Following this, she focused the students' attention on the learning objectives of the lesson at hand based on the objectives specified in the coursebook. Indeed, she seemed to attach particular importance to learning objectives as she frequently mentioned them at different stages throughout each lesson. In all of the lessons observed, the main areas of teaching and learning focus in the book were listening skills and vocabulary due to the structure and organization of the units. Thus, the learning objectives Sarah emphasized were mainly those that were related to listening skills. She frequently reminded the students about the target skills and made brief explanations about what they involved, drawing on the explanations provided in the book. Additionally, although it was not an area of focus in the coursebook during the lessons observed, Sarah also drew attention to pronunciation and taught the students practical strategies for correct pronunciation. In terms of vocabulary, it appeared to be one of the main areas of teaching focus, which was in line with the coursebook. Sarah taught the target vocabulary specified in the book explicitly through definitions, synonyms, antonyms and by modelling how to use the words, with a special focus on the students' own culture. As a matter of fact, most of the teacher

explanations were aimed at teaching vocabulary and listening and pronunciation skills. Activities for student practice regarding these main areas of teaching and learning focus mainly consisted of the completion of the exercises in the coursebook. For the student practice of the target vocabulary, Sarah additionally asked the students to make their own sentences for homework which would be reflective of their own experiences.

Sarah's effort to make learning relevant to students' own lives was also noticeable in the way she led the whole-class discussions, which she held at different stages of each lesson based on the discussion questions provided in the book. During these discussions, she clarified the discussion questions by giving examples from Turkish culture and mostly kept the discussions focused on Turkish culture. Also, she directed open-ended follow-up questions to both individual students and the whole class in order to elicit more details about their own opinions and experiences. In addition to whole-class discussions, she also instructed the students to have pair or group discussions at different stages in each lesson, guided by the instructions in the coursebook (i.e., "discuss the questions in a group"). As the students were having discussions, she carefully monitored each group apparently to make sure that they were having the discussions in English and to answer the questions of individual students.

In terms of the types of feedback she gave, she occasionally gave positive evaluative feedback and constant immediate oral corrective feedback throughout the lessons. With regard to her error correction style, she corrected all types of errors, at times implicitly and at others explicitly, including grammar errors.

4.3.2.2. Results from the Content Analysis of the Classroom Observation Data

The content analysis of the field notes and the transcript of the audio-recordings yielded two main categories and showed that Sarah's teaching displayed both characteristics of CLT and traditional approaches (see Table 8). The key aspects of her teaching such as explicit teaching of sub-skills, efforts to make learning relevant to the students' own lives and focus on student collaboration were in line with CLT whereas her apparent focus on accuracy during class discussions was suggestive of a traditional approach. Another important aspect of her teaching was that she taught vocabulary items explicitly.

Regardless of the approach she apparently adopted, when compared against her beliefs previously identified, on an observable level, her teaching methodology showed

no particular inconsistencies. As was found preciously, she encouraged group work, was careful about articulating learning objectives and gave constant corrective feedback. Also, her efforts to make learning relevant to the students' lives on both a classroom and an individual level could be interpreted as reflective of her belief that a teacher should recognize that students are individuals. However, although her methodology did not contradict her identified beliefs, evidently, in none of the lessons observed did she use any forms of media other than the coursebook audio material for the listening activities in the book; nor did she use any activities that could particularly be described as activities for promoting learning by doing.

These differences identified between Sarah's pedagogical beliefs and her actual classroom practices were noted to be brought up in the post-observation interviews.

Table 8.

Sarah-Results from the content analysis of the classroom observation data

Code	Example/Quote	Category
Explicit teaching of sub-skills	"So, today, we're going to listen for key words and key phrases." "Listen for key words. This will help you find the main idea. "If they're repeating the word, you can tell that it's a key word because they're using it more than one time." "If you find a word difficult to pronounce, just break it down, OK? Re-quire-ments." "Break it down. Ad-ver-tise-ment."	
Making learning relevant	"OK…For example, in Turkey, they ask about your age sometimes. They ask if you're married or single. What other questions do they ask?" "Say, for example, I come to your house and I walk into your house with my boots on. Would that be OK? Yes, I should take my boots off. This is a *custom* in Turkey."	**(1) Combination of CLT, with a focus on explicit skills instruction, and a traditional approach**
Student collaboration	Group and pair work: Asks students to think in groups of 3 interesting jobs and decide what skills are required for these jobs Asks students to discuss in pairs and decide how Turkey is different from other countries	
Focus on accuracy	Explicit correction of errors during class discussions: Student: "My resume." Teacher: "It's resumé. Not resume."	

Table 8 (continued)

English definitions	"What's advertising? To show a product, to show something they want you to buy."	
Synonyms	"Steps? Steps? Stages."	**(2) Explicit teaching of vocabulary**
Antonyms	"Peace? War, fighting is the opposite."	
Teacher modelling of new vocabulary	"Assistant? I wish I had an assistant, for example, to help me."	
Student practice of vocabulary	Fill-in-the-gaps exercises in the book	

4.3.3. Post-observation Interview Results

The post-observation interview with Sarah lasted for 60 minutes and took place in a comfortable spot in the university yard. In addition to the original five main interview questions, she was asked follow-up and probe questions depending on the direction of the conversation and the topics that emerged during the interview (see Appendix 3). She was also asked questions about the possible reasons for the identified differences between her beliefs and her teaching methods.

The content analysis of the interview transcript yielded nine main categories (see Table 9).

Sarah's responses to the interview questions indicated that she considered language skills to be too closely linked to each other to be separated. However, she also underscored the importance of different skills for different purposes, explaining that different skills should be given more weight depending on the learner's end goal. She held that listening, reading and writing were more important for academic purposes, whereas for personal purposes, listening and speaking would have greater importance. She also placed a particular emphasis on the importance of pronunciation and vocabulary knowledge in getting one's meaning across accurately. In terms of grammar, she seemed to consider it as a sub-skill rather than "a separate area on its own," and she did not think it played a crucial role in communication:

> *A person can make themselves understood even if the structure is not right.*
> *It (structure) is less important than the others to be understood and to*
> *understand.*

In fact, she believed grammar was the least important aspect of language:

> *I just mean, speaking from, like, a communicative aspect, then, you know,*
> *structure is the least important.*

Interestingly enough, however, when she was asked what aspects of the language she focused on the most in her daily practice, she listed grammar among the main areas of her teaching focus in addition to pronunciation and vocabulary. As to her teaching techniques, she mentioned giving constant feedback for pronunciation and grammar mistakes. Her description of her vocabulary teaching techniques suggested that she taught vocabulary explicitly with a special focus on student internalization of target vocabulary through personalized practice.

Her response to the question related to the assessment objectives of the exams mainly revolved around the use of different assessment tools for different assessment objectives, which she seemed to be content with. She believed the pop quizzes assessed all language skills in written form, the monthly tests assessed listening comprehension and the performance assessments assessed the students' actual speaking ability. She did not think there should be a speaking section on the pop-quizzes or the monthly tests, claiming that the students were sufficiently assessed on their speaking skills through presentations as it was.

She added, however, that although she did not think it was necessary to add a speaking section to the pop-quizzes and the monthly tests, she would like to add more speaking activities to do in class, complaining that there was too much focus on reading and writing in the book rather than speaking. She elaborated that she would like to take things from the students' real lives and put them into practice because she did not think the coursebook content was relevant to the students' lives:

> *In the books, it's all sort of hypothetical or there's things they can't relate*
> *to or they don't do, but they have to study these because it's in the book.*

According to Sarah, the exams were the primary factor that limited her freedom as a teacher and obviously, the book had to be covered because of the exams:

You are restricted. You <u>have to</u> do this because you know it's gonna be in the quiz, in the exam. You can't deviate from it very much. Even if you want to, even if the students are showing interest in something else, you can't pursue it too much. We're working within the constraints of the tests and exams and what we can teach.

The above excerpts strongly suggest negative washback from the exams on teaching content. She did not relate her focus on grammar errors to the exams, however:

If I spot an error, I always focus on it. To help the students use their grammar correctly.

In terms of how she would teach if she were not restricted, Sarah stated that she would like to use more mixed media and to encourage more group work and role playing activities in the classroom, which would provide more opportunities for students to practice what they had learned. Upon further probing, she explained:

How do you do it here? Because of the syllabus, because of the time schedule and the material.

It appeared that she considered learning by doing as student practice through an authentic context and role-play and group activities, and based on this definition, it appeared that the exams also exerted negative washback on Sarah's teaching methodology by causing her to devote less class time to group work and role playing activities as well as to the use of mixed media than she would like to due to the syllabus, pacing schedule and the material.

Overall, Sarah considered the exams as the primary factor that limited her freedom, and she ascribed the identified the differences between her beliefs and actual classroom practices mainly to the exams. However, she seemed to believe that the problem mainly arose from the syllabus, time constraints due to the intense pacing schedule and the material rather than the exams themselves.

When she was asked what would happen if she fell behind the pacing schedule and the topics she had not taught came up on the exam, she emphatically replied:

I'd be devastated. Because I would feel that I've let them (students) down because I know these points were important to the students. And if something came up on the pop quiz that I haven't covered, I would really feel so bad about it. Because you know, how can they answer the questions if they haven't studied it? I'd really feel terrible, really.

Thus, it seemed that the underlying cause of this concern over time pressures was her strong sense of accountability to the students.

Another finding that emerged from the interview concerns the role of teaching materials in Sarah's teaching. Throughout the interview, Sarah was noticeably focused on teaching material and the importance of using material with relevant content. In fact each time she mentioned teaching techniques, she made a reference to teaching material. Thus, it seemed as if material was of particular importance to Sarah. The following excerpt clearly suggests a significant role of teaching material in Sarah's teaching style:

Like I said….I don't think I'd have to change my teaching style so much because it would change with the material. Depending on what I'm using, my style changes.

Similarly, the following excerpt points to the role of coursebooks:

Because I haven't received any formal training in ELT, at the beginning of my career, I was at a loss how to teach…. So, I just went along with the coursebooks….And the different techniques from those coursebooks had considerable influence on the formation of the basis of my teaching style.

This excerpt clearly shows that Sarah relied on coursebooks to guide her teaching at the beginning of her career due to lack of formal training in ELT, and it also indicates that the coursebooks she used at the time played a major educational role in the development of her teaching style.

Taken together, these excerpts suggest that teaching materials play an important role in how Sarah teaches and thus have the potential to influence her teaching.

Table 9.

Sarah- Results from the content analysis of the interview transcript

Code	Example/Quote	Category
Importance of listening, reading and writing for academic purposes	"I'd suppose if it's for academic reasons, I'd say listening, reading and writing would be the main areas. But… for academic use."	**(1) Importance of different language skills for different purposes**
Importance of listening and speaking for personal use	"If it's for personal use, I'd say speaking and listening would be more important."	
Importance of pronunciation and vocabulary for communicative purposes	"Pronunciation is important because mispronunciation of words may cause misunderstandings in real-life situations." "Pronunciation could cause big problems. And a lack of vocabulary -the student can't express themselves."	
Grammar as a sub-skill	"Grammar is involved in all of those areas (speaking, listening, reading, writing) anyway. It's not so much a separate area on its own."	**(2) The limited role of grammar in communication**
Grammar less important than other skills	"It (grammar) is not as important as the others because you can still get the general meaning of what the student is trying to say even if the sentence isn't structured correctly." "…because a person can make themselves understood even if the structure is not right." "It (structure) is less important than the others to be understood and to understand."	
Least important aspect for communication	"I just mean, speaking from, like, a communicative aspect, then, you know, structure is the least important."	
Vocabulary lists in the book	"We go through the list in the book."	**(3) Explicit teaching of vocabulary with a focus on personalized practice**
Examples from the book	"There are examples in the book."	
Teacher modelling	"I give examples, also."	
Student example sentences for target vocabulary	"As long as I know that they can use that word correctly and understand the definition of the word and they can use it in the correct way, so, then I understand that they do understand the vocabulary."	

Table 9 (continued)

Code	Example/Quote	Category
Relevance to students' real life	"I would like to take things from the aspects of their real life." "I would like to take things from the aspects of their real life-their social life maybe."	**(4) Relevant Content**
Irrelevant content in the books	"In the books, it's all sort of hypothetical or there's things they can't relate to or they don't do…."	
Constant feedback for pronunciation mistakes	"Pronunciation is a constant. It's just a constant automatic feedback thing for me."	**(5) Constant corrective feedback for grammar and pronunciation errors**
Constant feedback for grammar mistakes	"I don't really teach grammar, but if I ever spot an error, I will focus on it.	
Assessing grammar Assessing speaking Assessing vocabulary Assessing listening	"A quarter of the quizzes-grammar. A quarter- skills…speaking skills, general skills. A quarter vocabulary. And then, a quarter listening." "For the monthly exams, they only do listening. There's no speaking section. Just listening comprehension. That's all."	**(6) Different assessment tools for different assessment objectives**
Assessing actual speaking ability	"They have presentations in every unit where they have to speak."	
Written assessment / Oral assessment	"In their quizzes, it's in written form; but in their presentations, it's verbal.	
Sufficient assessment of speaking skills	"I think it's OK. Because they do so many presentations ….I don't think it's necessary to give a speaking exam in the monthly."	
Focus on exam content	"but they have to study these because it's in the book." "You are restricted. You *have to* do this because you know it's gonna be in the quiz, in the exam. You can't deviate from it very much. Even if you want to, even if the students are showing interest in something else, you can't pursue it too much. We're working within the constraints of the tests and exams and what we can teach."	**(7) Washback on teaching**
Less class time devoted to group work / role playing activities / mixed media than she would like to	"I would encourage more group work and role play activities….I'd use more mixed media…."	

Table 9 (continued)

Code	Example/Quote	Category
Time constraints due to the intense pacing schedule	"There's too much to do in the time that we have." "In the beginning, I was trying to fit in extra things into the classroom, but it was causing us problems and we were falling behind the syllabus."	
Accountability to students	"I'd be devastated. Because I would feel that I've let them down because I know these points were important to the students. And if something came up on the pop quiz that I haven't covered, I would really feel so bad about it. Because you know, how can they answer the questions if they haven't studied it? I'd really feel terrible, really."	**(8) Causes of washback**
Material	How do you do it here? Because of the syllabus, time schedule and the material.	
Changes in style depending on material	"Depending on what I'm using, my style changes."	**(9) Influence of teaching material on her teaching style**
Role of the coursebook in the development of her teaching	"And the different techniques from those coursebooks had considerable influence on the formation of the basis of my teaching style."	

4.3.4. Summary of Relevant Findings for Sarah

A comparison of the results from the cluster analysis and those from the classroom observations showed that on an observable level, the key characteristics of Sarah's teaching methodology were mostly consistent with her beliefs. In line with her identified beliefs, she seemed to attach particular importance to the articulation of goals, giving prompt feedback and group work. Also, her focus on Turkish culture and on the experiences and opinions of individual students as well as her attentiveness to individual students seemed to be consistent with her beliefs regarding the recognition of students as individuals and relating new concepts to previous ones. However, there were also two differences between her identified beliefs and her actual classroom practices, which were lack of use of media and activities that encourage learning by doing.

The interview results indicated that the absence of these techniques from her methodology was a negative washback effect. Similarly, the exams also seemed to exert

negative, and apparently much stronger, washback on teaching content. Sarah did not think the coursebook content was appropriate for the students in terms of its relevance to their lives, and she strongly believed it would be much more beneficial for the students to learn through content relevant to their own lives; however, she had to prioritize the coursebook, regardless, due to the exams.

Despite these negative washback effects, Sarah did not seem to view the exams in a negative light. Overall, she seemed to consider the system as made up of different assessment tools aimed at different assessment objectives, which she seemed to be content with. Thus, it appeared that the negative washback effects were not a direct consequence of the exams, but rather of some other factors, which consisted of time constraints, the intense pacing schedule, the syllabus, and the material. It was also revealed through the interview that she felt a markedly strong sense of accountability to her students. The interview results also clearly indicated that the intense pacing schedule, which also functioned as a syllabus, was a major cause of time constraints; however, it seemed that the underlying cause of Sarah's concern over time was the strong sense of accountability she felt to her students. With respect to teaching material, it appeared as if teaching material was of particular importance to Sarah. She stated that her teaching style changed with the material. Also, it was revealed that Sarah relied on coursebooks to guide her teaching at the beginning of her career due to lack of training in ELT, which suggested an educational role of coursebooks in her career as a teacher. Based on these findings, it was inferred that teaching materials might have the potential to influence Sarah's teaching style.

4.4. Case 3- Jason

Jason is a Turkish-speaker teacher of English with a total of five years of teaching experience. At the time of the observations, he had been working at the institution for three years and six months, and was in his fourth academic year. He was teaching Coursebook to two different classes at the beginner level and ESP to international finance students, his total number of teaching hours amounting to 20 hours per week. He was not a coordinator. The observations took place in one of his Coursebook classes.

Table 10 presents the general characteristics of Jason along with the courses and levels he was teaching at the time of the observations.

Table 10.

Jason- General characteristics and the courses/levels he was teaching at the time of the observations

Native language	Turkish
Formal Training in ELT	Pedagogical Formation Certificate
Total years of teaching experience	5
Length of employment at the institution	In his 4th academic year
Courses/Levels he was teaching	Coursebook/ Beginner & ESP
Courses/Levels observed	Coursebook/ Beginner
Total number of teaching hours per week	20
Coordinator	No

4.4.1. Results from the Cluster Analysis of the Rep Grid Data

Jason generated a total of 14 constructs. The cluster analysis of these constructs yielded 2 distinct clusters (see Figure 3). The first cluster consists of two constructs, "good time management" versus "poor time management" and "uses audio-visual aids" versus "doesn't use audio-visual aids," which are matched at over 85%. The other cluster contains 10 constructs, eight of which formed four tight pairs. Among these pairs, "expertise" versus "uncertainty about subject knowledge" and "precise communication" versus "lack of communication," and "constructive feedback" versus "poor feedback" and "speaks clearly" versus "doesn't speak clearly" match at about 90%. The other two pairs, "patience" versus "impatience" and "an appropriate sense of humor" versus "too serious," and "empathy" versus "no empathy" and "open- minded" versus "narrow-minded" are matched at over 90%. Within the same cluster, "organized" versus "disorganized" stands as an isolated construct and the construct, "engagement" versus "static" is linked to the pair "empathy" versus "no empathy" and "open- minded" versus "narrow-minded," at about 90%. The constructs "uses various teaching methods" versus "uses traditional methods" and "improvisation" versus "rigidity" stand alone as isolated constructs.

Figure 3. *Cluster analysis of Jason's constructs about ideal teaching methods*

The first cluster in the FOCUSed grid suggests that Jason believes an effective teacher has good time management skills and uses audio-visual aids, and the relationship between these constructs suggests that he believes the use of audio-visual aids requires efficient use of time.

Taking into account the connections between the constructs in the larger cluster, it could be inferred that Jason believes effective instruction requires the teacher to communicate effectively, which requires expertise; to be organized; to provide constructive feedback, for which it is necessary to speak clearly; and to show patience for the students, which can be achieved through an appropriate sense of humor. Also, he seems to believe that open-mindedness helps a teacher empathize with his students, which in turn enables him to be able to engage with them actively.

The construct, "uses various teaching methods" versus "uses traditional methods" stands alone; however, it is noteworthy in its own right in that "uses various teaching methods" is contrasted specifically with "uses traditional methods," which implies that according to Jason, a teacher should ideally include different teaching methods in his teaching but not traditional methods.

4.4.2. Observation Results

4.4.2.1. Overview of Jason's Lessons

The classroom observations of Jason were particularly interesting in that there were significant differences between the first two lessons and the next two lessons observed. They differed not only in terms of the areas of teaching focus and the

teaching materials used but also in terms of the teacher's overall attitude and his teaching methodology. Indeed, Jason showed such distinctly different teaching behaviors in the first and the next two lessons observed that each two-lesson block of observation almost felt like the observation of a different teacher.

In the first two lessons observed, the only focus of teaching was grammar and more specifically the passive voice, and the teaching materials used were two syllabus-based, form-focused worksheets apparently aimed at helping students practice forming sentences and questions in the passive voice. Neither the worksheets nor the lessons in general included any meaningful or communicative activities. Also, there was no student-to-student interaction, and the learning activities were limited to listening to teacher explanations and doing the exercises in the worksheets. In terms of teaching activities, Jason's teaching was characterized by teaching grammar deductively through formulas and frequent use of grammatical terminology, which was also core to the type of corrective feedback he gave the students. He typically corrected student errors immediately and explicitly, allowing the students little to no wait-time at all. In terms of vocabulary teaching, he taught vocabulary through translations for the most part and only when the students asked what a particular word in the worksheet meant. Also, his authoritative attitude was palpable, and the lessons were strictly controlled by the teacher. Despite the strict teacher control, however, Jason looked a little detached during the lessons. Overall, to an outsider, he did not seem to be particularly enjoying the worksheets.

In the next two lessons observed, the teaching material was the coursebook, and both the teaching content of the lessons and the activities used were mainly based on the content and activities provided in the book. In terms of Jason's teaching style, as mentioned before, Jason, surprisingly, seemed like a completely different teacher in these two lessons. In contrast to his form-focused traditional approach observed during the first round of observations, his teaching style in the next two lessons observed was characterized by an interactive style focused on fluency and authentic communication.

He typically held whole-class discussions, during which he asked individual students open-ended questions probing into their personal experiences and opinions, apparently in an attempt to make the new information relevant to their own lives. He also encouraged the students to improvise the role-playing activities in the book, which he made a special effort to link to the relevant listening activities. With respect to vocabulary teaching, he mostly gave simple English definitions for the target

vocabulary and occasionally provided one or two-word translations. As for his error correction style, he tended to correct vocabulary errors explicitly whereas he occasionally corrected grammar and pronunciation errors implicitly during whole-class discussions. During role-playing activities, however, he did not correct any mistakes or errors. It is also worth noting that he seemed much more energetic and enthusiastic in these lessons compared to the previous two lessons observed.

4.4.2.2. Results from the Content Analysis of the Classroom Observation Data

The results from the content analysis generated three main categories, and showed that Jason's teaching in the first two lessons observed was characterized by a form-focused traditional approach and that he was mostly inconsistent with his previously identified beliefs during the lessons in question (see Table 11). In contrast to his beliefs, in the first two lessons observed, he mostly used traditional methods; gave no feedback that could specifically be described as constructive; looked quite authoritative and serious throughout the lessons rather than humorous; and allowed very little to no wait-time, which suggested that he did not have much patience for student mistakes and errors. Also, he did not use any audio-visual aids, and overall, his engagement with the students seemed to be limited to correcting their mistakes.

However, the results also showed that on an observable level, his teaching behaviors in the third and the fourth lessons observed were mostly consistent with his beliefs. In those lessons, his teaching was characterized by meaningful activities such as improvised role-playing activities, the integration of listening and speaking skills, focus on fluency and on students' own experiences, which could be classified as CLT focused on speaking. Another key characteristic of his teaching was explicit teaching of vocabulary. Unlike in the previous two lessons observed, he gave English definitions and examples in addition to providing translations. Also, the students practiced the target vocabulary through the gap-fill exercises in the book.

These identified inconsistencies and differences were noted to be brought up in the post-observation interview.

Table 11.

Jason- Results from the content analysis of the classroom observation data

In the first two lessons observed- Teaching material: Worksheet on the passive voice		
Code	**Example/Quote**	**Category**
Focus on grammar only	Form-focused grammar exercises about the passive voice for 2 teaching hours No meaningful or communicative activities	
Deductive teaching of grammar	First writes the rule, then gives an example and asks students to do the worksheet	
Formulaic explanations	"We use the subject first and then.. auxiliary- am/is/are/was/were and V3. V3, right? The past participle."	
Use of grammatical terminology	"First, we use the question word and we use the auxiliary, right? And then, subject comes. And the last one is V3."	
Immediate and oral explicit/ metalinguistic corrective feedback	Typically gives explicit and metalinguistic corrective feedback: Student: "The man is seen on the stairs." (as the answer to the question "where was the man seen?") Teacher: "It's in the past, so it's was. The man *was* seen… "V3? Where is the verb 3?"	**(1) Traditional approach to teaching**
Teaching of vocabulary on an ad hoc basis through translations	"Hijack, kaçırmak in Turkish." "Compose means, in Turkish, bestelemek."	
No student-to -student interaction	Only teacher initiated teacher-to-student interaction	
Short-wait time	Leans toward students who give the correct answers or he himself provides the correct answer immediately	
Authoritative	Teacher makes all the decisions such as explaining problem areas and calling on students despite the presence of volunteering students	

Table 11 (continued)

Code	Example/Quote	Category
Extended teacher-student interaction during whole-class discussions	Teacher: Which one (hotel) would you recommend to a visitor?" Student: " Hilton." Teacher: "Hilton? Why?" Student: "Hilton is in the city center." Teacher: "What about the public transport? Is it easy to get there?" Student: "Yes. You can walk everywhere" Teacher: "Yeah.. You can walk...But there are a lot of hotels in the city center… Why Hilton?"	
Focus on students' own experiences	"Anyone who met an old friend on Facebook? Here's a personal question: What do you talk about with an old friend? Oh.. like.. "Do you remember that girl? That boy… He was in love with you?"	**(2) CLT focused on speaking**
Improvised role-play activities	Before the role-playing activities, he especially asks students not to write anything down, and first he models the activity with the students and then has the students improvise on the conversations in the book.	
Integration of listening and speaking skills	"Let's listen to the conversation again and then we'll do something similar (plays the recording).	
Focus on fluency	Implicit corrective feedback during whole-class discussions: Student: "Because good view." Teacher: "Good view. Yeah… because it has a good view." Does not correct any errors as students role play	
English definitions for the target vocabulary	"Laundry is the place where people do the washing. Dirty clothes go there and then they come clean."	
Turkish translations for the target vocabulary	"So, it (check in) means giriş yapmak."	**(3) Explicit teaching of vocabulary**
Explicit error correction	"We can't say polluted. It's dirty, we say."	
Gap-filling exercises for vocabulary	Exercises in the coursebook	

In the next two lessons observed- Teaching Material: Coursebook

4.4.3. Post-observation Interview Results

The post-observation interview with Jason lasted for 64 minutes and 44 seconds and took place in a comfortable spot in the university yard. At the participant's request, the full recording was shared with him the following day. In addition to the original five main interview questions, he was asked follow-up and probe questions depending on the direction of the conversation and the topics that emerged during the interview (see Appendix 3). He was also asked questions about the possible reasons for the inconsistencies between his beliefs and the teaching methods he used in the first two lessons observed. The content analysis of the interview transcript yielded nine main categories (see Table 12).

Jason's responses to the interview questions showed that he considered language skills as an inseparable set of interrelated skills, among which he enjoyed speaking the most. Actually, he seemed to have a particular interest in speaking as he often turned the conversation back to speaking skills. Overall, it appeared that of all the four language skills, speaking was the one he felt the most comfortable about. However, when he was asked what he focused on the most in his daily practice as a teacher, he replied:

Grammar, I think. Grammar is probably what I focus on the most, nevertheless.
He went on to explain that although he did not believe it was the most important aspect of the language, he focused on grammar the most because he believed it was the most difficult aspect of the language. He also mentioned his own learning experiences as one of the underlying reasons for his focus on grammar, which was a topic he brought up frequently throughout the interview:

> *And of course…because this is how we were taught. It's not easy to break*
> *with tradition. No matter how open-minded you try to be to new ideas and*
> *to keep up to date, I think the teacher's own learning experiences have an*
> *important influence on their teaching techniques.*

His description of how he taught grammar was suggestive of explicit, deductive teaching of grammar with a special focus on simplicity. As for how he taught vocabulary, it appeared that he believed he taught target vocabulary explicitly through translations and examples and with a particular focus on the meanings and correct usage of new words.

With regard to his views of the exams, Jason seemed to believe that the exams had three main functions: to test grammar and vocabulary, to help both the students and teachers focus on a clear goal to achieve and to provide concrete feedback on student learning. He also believed that there was a good alignment between what was taught in class and what was tested on the exams in terms of grammar; however, in terms of speaking skills, he did not think the exams efficiently assessed what was covered in class, and thus he strongly believed speaking should be assessed on the monthly tests as well:

> *But still, the exams don't really assess all those things from the book we do in class. I think it's definitely necessary to include speaking in the monthly tests …. It's almost impossible to do that on the pop quizzes, but I think it (speaking) should definitely be included in the monthly tests…. It's a major weakness of those exams.*

Apparently, for Jason, the reason for the necessity of including speaking in the monthly tests was related to the motivational influence of the exams on student learning. He seemed to believe that the exams motivated the students by making them focus on a clear goal and thus contributed to learning. However, at the same time, he believed the exams sometimes demotivated the students by discouraging them and causing them to lose interest:

> *but sometimes it has the obverse effect: the student does badly on an exam and thinks 'I can't do it anyway,' and then he lets things slide.*
> *…and, it's like…They take the quiz. They're like 'the quiz has come. OK. I was waiting for this. Now that I've done the quiz, the danger is over, and I can go home now.' And then, they don't attend the (next) lesson.*

He ascribed such demotivation mainly to frequent testing and the excessive number of exams:

> *I think in our school, there is a bombardment of quizzes and exams. And you know what? After a certain point, that causes the students to become*

numb. You know why? Because after you've slapped a man five times, it doesn't really hurt the sixth time…. They become numb to everything.

He further elaborated by means of a telling memory:

I remember one particular day. In the first hour, they had a vocabulary exam; in the third hour, we gave them a pop quiz; and then, in the afternoon, they did a presentation in the skills lessons. And, the students' eyelids were drooping in the end.

Obviously, Jason believed, the exams both motivated and demotivated the students, which are suggestive of both positive and negative washback from the exams on students and student learning. Like Emma, he seemed to consider frequent testing and the excessive number of exams to be the underlying cause of student demotivation, that is, negative washback from the exams. Nevertheless, Jason did not mention anything that could directly be related to washback from the exams on his teaching until toward the end of the interview. When he was asked about the potential reasons for the differences between the first two and the next two lessons observed, he lamented that the worksheet was rather mechanical and unnatural:

Because the worksheet was so mechanical. You're like a machine….You're like a robot. It's not natural. You only bring a couple of words together…. Does it help me communicate with anyone? No…. But, in the book, you read a reading passage from some magazine. You talk about it. There is life in it…. Life stops when you're doing those worksheets. It's like everybody's tightening a screw in a factory. This is how it feels to me.

When he was asked why he used those worksheets if he disliked them so much, he replied:

Because I have to. Not doing the worksheets? (Laughs incredulously). I'd get fired…. If I didn't do those worksheets, my students would complain to the administration about me when those questions come up on the exams.

These two responses when considered in conjunction with the observation results clearly point to strong and material-induced negative washback on his teaching methodology as well as on teaching content as a result of a strong sense of accountability to the administration and incompatibility between the style of the worksheet and his beliefs.

The interview also revealed some other beliefs Jason had about ideal teaching methods, which he had not stated in the previous stages of the study. He frequently mentioned the importance of authenticity in teaching and learning English, and explained that he would like to create a natural environment for the students where they would have to use English, which would encourage learning by doing. He listed a number of personal reasons such as lack of time and personal priorities as the primary reasons for not implementing these ideas in his daily practice.

Table 12.

Jason- Results from the content analysis of the interview transcript

Code	Example/Quote	Category
Favorite skill	"Personally, I enjoy speaking the most."	**(1) Area he feels most comfortable about**
Confidence in his speaking ability	"I think speaking is my strongest side." "I've never failed an oral English exam."	
Grammar as the most difficult aspect of the language	"But I think we focus on it the most because it's the most difficult aspect of the language."	**(2) Reasons for focus on grammar**
Teachers' own learning experiences	"And of course because that's how we were taught. It's not easy to break with tradition."	
Teacher explanations of grammar rules	"I explain the logic behind it."	**(3) Deductive, explicit teaching of grammar focused on simplicity**
Level-appropriate teacher explanations of grammar	"I usually do it in Turkish depending on the students' background knowledge and level of proficiency."	
Examples provided by teacher	"I write examples on the board and I ask them to take notes."	
Simplicity	"I try to do that as plainly and simply as possible."	

Table 12 (continued)

Code	Example/Quote	Category
Word meanings	"We talk about the meanings first"	
Turkish translations and focus on word usage	"I tell them the Turkish equivalents of the words and show them how the words are used.	**(4) Explicit teaching of vocabulary**
Teacher modelling	"Then, I give some examples."	
Real-life situation	"I would bring native speakers of English and learners of English together and make them talk to each other".	
Direct experience	"And the learners would consist of people with different native languages-so that they would all have to speak English". "I would create a natural environment….so that they would be immersed in the natural flow of the language."	**(5) Experiential Learning**
Teacher support	"And I would supervise them."	
Integration of academic knowledge and direct experience	"There would be a classroom environment as well….I'd like the two environments to work systematically."	
Assessment of grammar /Assessment of vocabulary	"It was only grammar until this year. This year we have also started to test vocabulary."	
A goal to achieve	"There should be some sort of assessment because it sets a clear goal for both the teacher and the student to achieve."	**(6) Functions of the exams**
Feedback	"It also provides some concrete feedback."	
Student motivation	They (the exams) definitely contribute to learning because they help the students focus on a goal….and provide motivation."	**(7) Washback on students and student learning**
Student demotivation	"…but sometimes it has the obverse effect: the student does badly and thinks 'I can't do it anyway,' and then he let things slide."	

Table 12 (continued)

Code	Example/Quote	Category
Focus on exam content	"Because I have to" (I'd get fired. If I didn't do those worksheets, my students complain to the administration about me when those questions come up on the exams.")	
Changes in teaching style depending on the material	"Because the worksheet was so mechanical…. You only bring a couple of words together…. All you do is to say, you're like, "this is correct, this incorrect, this is it and it's that'…. But, in the book, you read a passage from some magazine. You talk about it."	**(8) Washback on teaching**
Too many exams	""In our school, the students are bombarded with quizzes and exams." "There are so many exams that after a certain point, they become impervious to the pain."	
Accountability to the administration	"I'd get fired….If I didn't do those worksheets, my students complain to the administration about me when those questions come up on the exams."	**(9) Causes of washback**
Incompatibility between teaching material and his preferred methods of teaching	"Because the worksheet was so mechanical. You're like a machine. You're like a robot. It's not natural….Does it help me communicate with anyone? No…. But, in the book, you read a reading passage from some magazine. You talk about it. There is life in it. Life stops when you're doing those worksheets. It's like everybody's tightening a screw in a factory. This is how it feels to me."	

4.4.4. Summary of Relevant Findings for Jason

A comparison of the findings from the cluster analysis and those from the classroom observations revealed major inconsistencies between Jason's previously identified beliefs and his teaching behaviors in the first two lessons observed. In those lessons, Jason's teaching was in marked contrast to his beliefs in that his teaching patterns mainly classified as a traditional approach to language teaching; he was very serious rather than humorous; he provided no constructive feedback; he did not use any audio-visual aids; and his engagement with the students was limited to correcting their mistakes. Also, he allowed the students little, if any, wait- time, which suggested that he

did not have much patience for student errors and mistakes. However, in the third and the fourth lessons observed, he was mostly consistent with his beliefs. Indeed, his teaching was so different from how it was in the previous two lessons observed that he almost appeared to be a completely different teacher.

The interview additionally revealed other beliefs Jason had about ideal teaching methods which emphasized the encouragement of learning by doing. The primary reasons he listed for not implementing this into his daily practice were related to personal priorities. Another interesting finding regarding beliefs was his belief about the influence of teachers' own learning experiences on their teaching techniques. It was clear that Jason believed the way teachers are taught have a profound and long- lasting effect on the way they teach.

The interview also revealed that Jason believed the pop- quizzes and the monthly tests had three main functions: to test grammar and vocabulary, to help both the students and teachers focus on a clear goal to achieve, and to provide concrete feedback on student learning. Overall, he did not seem to be displeased with the design of the tests except that they did not assess speaking skills. He considered the lack of a speaking section on the monthly tests in particular as a major weakness of those exams.

In terms of washback effects, Jason believed the exams influenced student learning positively by motivating the students and at the same time negatively by causing demotivation and lack of student interest. Like Emma, he ascribed this negative effect to frequent testing and the excessive number of exams. As for washback from the exams on Jason's teaching, the complementary results pointed to strong and material-mediated negative washback resulting from incompatibility between the style of the teaching material and his beliefs in combination with a strong sense of accountability to the administration, which explained the inconsistencies between his beliefs and teaching practices in the first two lessons observed.

CHAPTER V

5. CONCLUSION

5.1. Introduction

This chapter summarizes the study and discusses the major findings that pertain to the research questions of the study. Then, it presents a discussion of the implications of the study and finally concludes with suggestions for future research.

5.2. Summary and Discussion

The aims of this study were twofold. The primary aim was to investigate the washback effects of the institutional progress tests on teachers' teaching methodologies in the English preparatory school of a university in Turkey. The secondary aim was to explore the role of the teachers' pedagogical beliefs and their perceptions of the tests in the mediation of washback. The research questions the study aimed to address were:

1) What are the washback effects of the institutional progress tests administered at the English preparatory school of a university in Turkey on teachers' teaching methodologies?

2) What role do the teachers' pedagogical beliefs and their perceptions of the progress tests play in the generation or inhibition of washback on their teaching methodologies?

Taking into account the potential role of contextual factors in the generation of washback, the study adopted a qualitative multiple case study approach. Multiple cases were used based on the understanding that "studying multiple units can provide better illumination" and help enhance transferability (Ary et al., 2010, p. 456; Merriam, 1998; Rowley, 2002; Shenton, 2004). The participants in the study were three teachers, who were not coordinators. Data collection methods consisted of the repertory grid technique, classroom observations and post-observation interviews with the teachers. The repertory grid technique was used to identify the teachers' pedagogical beliefs, classroom observations to identify whether the teachers' actual classroom practices corresponded to their pedagogical beliefs. The post-observation interviews were intended to complement the rep grid data by gathering further information about the teachers' beliefs about ideal teaching practices; to gather information about the

teachers' perceptions of the tests; and to help gain an in-depth understanding of why exactly the teachers did what they did in the classroom and to find out whether and how it was related to the exams. Detailed field notes were taken during the classroom observations, and all the lessons observed were audio-recorded. Also, the post-observation interviews were audio-recorded. Data collected through the repertory grid technique were analyzed using cluster analysis. For the analysis of classroom observations, firstly the audio-recordings were fully transcribed. Then, field notes and the transcriptions of the audio-recordings were analyzed as a single unit of text using content analysis. Similarly, for the analysis of the post-observation interviews, firstly the audio-recordings were transcribed, and then the transcriptions were analyzed using content analysis. A comparison of the results from the repertory grid data and those from the classroom observations revealed both consistencies and inconsistencies between the pedagogical beliefs and actual classroom practices of the teachers. For two of the teachers, the results showed that the inconsistencies were washback effects of the tests whereas in the other teacher's case, the post-observation interview revealed that the inconsistency was due to another belief that had not been identified in the previous stages of the study.

The details of major findings are discussed below in light of the research questions that the study aimed to address:

1) What are the washback effects of the institutional progress tests administered at the English preparatory school of a university in Turkey on teachers' teaching methodologies?

At the most general level, the primary washback effect of the tests was the prioritization of exam-related teaching materials. Both the results from the observational data and those from the post-observation interviews yielded strong evidence that all of the teachers prioritized exam-related materials such as the coursebook and supplementary worksheets. However, although such prioritization naturally resulted in an intense focus on and the prioritization of the content of the materials, that is exam content, apparently, the scope of the influence of the materials on teaching was not limited to teaching content. The results indicated that strict adherence to the teaching materials also led to washback on the teachers' teaching methodologies.

Overall, the washback effects of the tests on the teachers' teaching methodologies consisted of the elimination of the teacher's preferred approach, methods and/or techniques or decrease in the amount time allocated to certain techniques due to

the prioritization of the exam-related teaching materials. It was found that in their efforts to prioritize the teaching materials, the teachers opted either to eliminate their preferred approach, method and/or techniques from their teaching or to allocate less time to some techniques than they previously did or would ideally prefer to. In the case of Emma, it was found that she eliminated activities such as games, vocabulary gap-fill and sentence completion exercises and matching activities from her teaching in the second semester. She also reported limiting pair work and group work activities, which suggested that she devoted less time to those activities. The results indicated that Sarah also chose to devote less time to techniques that she would prefer to spend more time on such as role playing activities, group work and use of mixed media. As for Jason, it was revealed that he forwent his method of choice altogether in the face of incompatibility between his pedagogical beliefs and the style of the material and instead, employed a traditional approach and relevant techniques inconsistent with his beliefs.

Based on these findings, it is not easy to make a definite judgement about the overall value and intensity of the washback on the teachers' teaching methodologies, however. In terms of the value of the washback, the results indicated that the prioritization of the exam-related materials resulted in negative washback for all of the teachers: Jason adopted a traditional approach and used techniques inconsistent with his beliefs in the first two lessons observed, and apparently Emma and Sarah devoted less time to techniques that are more conducive to language learning such as pair work, group work, games and use of mixed media. On the other hand, however, all of the teachers seemed to adhere not only to the content of the coursebooks they used but also to the activities and the instructions provided in the books for one reason or another, which may have contributed to the generation or intensification of positive washback in the shape of increased use of class discussions, pair and group work activities. Also, Emma stated that she had to eliminate activities like vocabulary gap-fill and sentence completion exercises and matching activities; however, it could be argued that if the elimination of such relatively mechanical activities led Emma to focus more attention on the communicative activities provided in the coursebook, that could be regarded as a positive washback effect. On the other hand, it could also be a negative washback effect if she chose to eliminate these activities despite the specific needs of her students for more mechanical practice. In terms of the intensity of washback, there was strong evidence of strong negative washback on Jason's teaching methodology due to incompatibility between his beliefs and the style of the material; however, the study

found no direct observational evidence that supported the other two teachers' claims about negative effects of the tests on their methodologies. Although Emma did not use games, matching activities or vocabulary gap-fill and sentence completion exercises during the lessons observed, there is no knowing whether she really used those activities in the first semester; nor can this study provide direct evidence regarding whether she really devoted more time to pair/group work activities in the first semester. Similarly, although it is true that Sarah did not use any role-playing activities or mixed media during the lessons observed, it is not clear whether she would really use such techniques in the absence of the exams. Also, based on the findings of the study, it is not possible to make a judgment about how much more time she would really devote to group work if it were not for the exams. Thus, it is not clear how strong the washback really was on their methodologies.

It should be noted that the prioritization of the exam-related materials itself was a washback effect of the tests, which also gave rise to washback on the teachers' teaching methodologies. However, neither the tests themselves nor the prioritization of the materials was the underlying cause of washback from the tests.

2) What role do the teachers' pedagogical beliefs and their perceptions of the progress tests play in the generation or inhibition of washback on their teaching methodologies?

The results suggested a potential role of the teachers' pedagogical beliefs in the generation, intensification, mitigation and inhibition of washback from the tests on their teaching methodologies either directly or indirectly as a result of the interaction of their existing beliefs with exam-related teaching materials.

In Jason's case, the study found strong evidence that the level of compatibility between Jason's predominant pedagogical beliefs and the style of the materials he used was a major factor that contributed to the generation and inhibition of washback from the tests on his teaching methodology. The complimentary results convincingly showed that incompatibility between the style of the supplementary material he used in the first two lessons observed and his predominant pedagogical beliefs resulted in strong negative washback on his teaching methodology, which consisted of Jason's use of a traditional approach and relevant techniques inconsistent with his predominant beliefs. It was also found that the compatibility between the style of the coursebook and Jason's pedagogical beliefs might have played a role in the inhibition or mitigation of negative washback from the tests on his teaching. The post-observation interview results showed

that Jason considered the coursebook to be more compatible with his predominant beliefs. The interview also found that in terms of the assessment objectives of the tests, he held that the exams mainly aimed to assess grammar and vocabulary, and he considered that lack of a speaking section on the monthly tests especially was a major shortcoming of the progress tests administered at the institution. However, there was observational evidence that he focused on speaking and encouraged the students to improvise on the role-playing activities provided in the coursebook during the lessons he taught the coursebook, which suggests that based on his description of the assessment objectives of the tests, he focused on speaking, improvisation and role-playing activities in spite of the tests, and he might thus have inhibited or at least mitigated the potential negative washback from the tests.

For Emma, the results suggested that overall, her pedagogical beliefs might have played a more direct role in the intensification of both positive and negative washback from the tests compared to Jason's. Whereas Jason's teaching style seemed to vary with the material, and he focused on grammar only while teaching grammar-focused supplementary materials, Emma's focus on grammatical accuracy was manifest in all of the lessons she taught although other aspects of her teaching were in line with her belief that a teacher should ideally use CLT methods. The post-observation interview with Emma suggested that in addition to her belief that CLT is the best approach to language teaching, she might have an ingrained belief that grammatical accuracy is of particular importance, which appeared to be the reason for her particular focus on grammatical accuracy during all of the activities she held in class. In terms of her perceptions of the assessment objectives of the tests, she considered that the tests mainly tested grammar, language functions, vocabulary, listening and reading skills. Emma's description of the tests suggests that the tests consisted of elements reflective of both traditional language teaching and CLT. Based on this description and taking into account her belief regarding the importance of grammatical accuracy and her classroom practices, it could be speculated that the belief in question might have resulted in the intensification of any potential negative washback from the exams. On the other hand, however, her belief regarding the necessity of adopting a communicative approach might have helped intensify the potential positive washback that might result from the communicative aspects of the tests. It is also worth noting that Emma often referred to the coursebook during the post-observation interview as if to justify her teaching method and techniques, which suggested that the book itself might have played a role in the

consolidation of her beliefs regarding CLT, and that might be an additional factor that could potentially contribute to the intensification of potential positive washback from the tests.

As for Sarah, the complimentary results showed that she believed making learning relevant was of particular importance and that her actual classroom practices were consistent with that belief. When this is considered in conjunction with the finding that Sarah regarded the irrelevant content of the teaching material as a major cause of negative influence of the tests, it could be argued that her belief regarding making learning relevant might potentially mitigate negative washback from the tests on teaching content. The potential role of her pedagogical beliefs in the generation or inhibition of washback on her teaching methodology, however, was less clear than it was for the other two teachers mainly due to her intense focus on teaching content and teaching material rather than methodology. The post-observation interview indicated that teaching materials had a significant influence on her teaching style as was the case for Jason. However, a major difference between Sarah and Jason with respect to how their teaching styles were influenced by teaching materials was that although there was strong evidence that level of compatibility between his predominant beliefs and the style of the material was a determining factor for Jason that resulted in washback on his teaching, no such evidence was found for Sarah. Also, for Sarah, considering her intense focus on content and material throughout the interview to the extent that she seemed to regard teaching style as coterminous with teaching material and that the coursebooks she used at the beginning of her career played a significant role in the development of her teaching style as well as her strict adherence to the activities provided in the book during the lessons observed, it could tentatively be suggested that the influence of teaching materials on her teaching style could also partly be due to the unestablished nature of her pedagogical beliefs.

Although it could be argued that the teachers' pedagogical beliefs might have played a role in the generation of washback in the given context, the results indicate that the teachers' pedagogical beliefs per se were not the main cause of washback. It appeared that in fact the underlying cause of all washback was the teachers' perceptions of the tests in terms of their consequences. The results from the post-observation interviews clearly showed that all of the teachers felt a strong sense of accountability, Emma and Jason to the administration and Sarah to her students. Emma was positive that her students would complain to the administration if she fell behind the pacing

schedule and failed to teach all the topics that could come up on the pop quiz; Jason seemed convinced that he would be "fired" if he did not use the supplementary material because similar questions could come up on the pop quiz, and obviously deeply bothered even by the thought of it, Sarah stated that she would be devastated if she fell behind the pacing schedule and the topics she had not taught came up on the pop quiz because "these points were important for the students." These responses strongly suggest that the teachers perceived the stakes of each individual pop quiz to be high. Obviously, Emma and Jason believed the stakes were high for themselves even if not for the students and Sarah both for herself-probably with regard to her relationship with her students- and her students. This is quite surprising considering that in fact, each individual pop quiz carried a negligible amount of weight within the whole scoring system. Overall, this finding is of particular importance for two reasons: firstly, because it appears to be the main reason why the teachers prioritized the exam-related material and secondly, because it clearly shows that when institutionalized, even tests that are actually very low-stakes in nature can be perceived as high-stakes by teachers and might thus result in washback effects similar to those caused by any high-stakes test.

Another factor that seemed to play an important role in the generation of washback within the context where the study was conducted was time constraints due to the intense pacing schedule. These two themes emerged as recurrent themes during the interviews with Emma and Sarah, and both of the teachers complained that they did not have enough time to do all the activities they would like to due to the intense pacing schedule, which indicated that time constraints due to the intense pacing schedule contributed to the generation of washback in the shape of the elimination of certain activities and techniques and decrease in the amount of time allocated to some. However, although it is clear that time constraints were a major factor that contributed to the generation of washback, it is worth reiterating that the main underlying cause of washback seemed to be the teachers' strong sense of accountability. It appeared that the teachers chose to prioritize the exam-related material and content essentially because of a strong sense of accountability and because they prioritized the material and the content, they did not have enough time left to do the other activities they would prefer to.

Based on these findings, the mechanism of washback from the progress tests on the teachers' teaching methodologies could briefly be summarized as follows: The teachers' strong sense of accountability caused them to prioritize the exam-related

materials, which in turn, resulted in washback on teaching content and teaching methodology. The washback on teaching content was direct and clearly observable. However, on an observable level, washback on teaching methodology in large part seemed to be a result of the interaction of the teachers' existing beliefs with the style of the material although the results also suggested that both the teachers' pedagogical beliefs and the methodologies promoted by the materials, and especially those promoted by the coursebooks might have the potential to directly influence washback on teaching methodology. Another important factor that obviously played a role in the generation of washback was time constraints. It appeared that the teachers prioritized the exam-related materials because of a strong sense of accountability and because they prioritized those materials they did not have sufficient time to do the activities they would prefer to, which caused them to eliminate certain methods and techniques and/or decrease the amount of time they allocated to some activities. With regard to the specific roles of the different factors in the generation or inhibition of washback, it could be inferred that, within the specific context where the study was conducted, the teaching materials were not the cause but the source of washback; the teachers' sense of accountability was the main driver of washback; and time constraints, and the teachers' pedagogical beliefs when they contributed to the generation of washback acted as intensifiers of washback. Also, the teachers' pedagogical beliefs could be regarded as a buffer against washback in cases where they served to mitigate or inhibit potential negative washback.

In conclusion, the results of the study showed that the progress tests administered at the given institution exerted washback on both teaching content and the methodologies the teachers used. In line with previous research, the findings of the present study indicated that washback on teaching content was direct, strong and uniform whereas washback on the teachers' teaching methodologies occurred as a result of different factors, and both the value and the intensity of washback on these areas varied among the teachers (Badger & Yang, 2012; Cheng, 2004; Wall & Horak, 2011; Wang, 2010; Watanabe, 2004a; Watanabe, 2004b). The finding regarding the washback effects of the tests on teaching methodology is also consistent with the argument that tests may affect different teachers differently and to different degrees (Alderson & Wall, 1993; Watanabe, 2004a). The factors that seemed to play a role in the generation of washback on the teachers' methodologies broadly consisted of the teachers' perceptions of the tests, the interaction of their predominant beliefs with the style of the materials

they used and time constraints. Similar to the findings from the study by Wall and Horak (2011), the results also showed that the teachers' classroom practices were strongly guided by the coursebooks they used. Although it appeared that this was mainly due to compatibility between the styles of the coursebooks and the teachers' existing predominant beliefs, the results also suggested that coursebooks had the potential to influence the teachers' pedagogical beliefs. Also, overall findings suggested that the teachers' existing pedagogical beliefs had the potential to exert direct influence on the generation or inhibition of washback, which is in line with the findings from previous research (Wall & Horak, 2011; Wang, 2010; Watanabe, 2004b). The underlying cause of washback, however, appeared to be the teachers' strong sense of accountability. Surprisingly, the teachers' perceptions of the tests with regard to their consequences suggested that they viewed even each individual pop quiz as a high-stakes test. This supports the argument that "if students, teachers or administrators believe that the results of an examination are important, it matters very little whether this is really true or false" (Madaus, 1988, p. 35). When the whole context is taken into account, the results of the present study suggest that the participant teachers' surprisingly high levels of sense of accountability could well be due to the institution-specific factors such as efforts to standardize teaching and to control teacher behavior by means of teaching materials and exams. Considering that the teachers prioritized the exam-related materials because of a strong sense of accountability, which led them to eliminate more preferable methods or techniques from their teaching or to allocate less time to activities such as pair and group work due to time constraints, it appears that such efforts may give rise to negative washback. On the other hand, however, the results also suggest that standardization efforts do work indeed and may have positive effects on teaching thanks to the coursebooks used, each of which apparently adopted a communicative approach to teaching. Although there were naturally differences between how the teachers taught, during the lessons when the teachers taught the coursebook, their overall teaching styles did not actually vary dramatically. As stated before, while teaching the coursebook, they not only adhered to the content but also for the most part implemented the activities in the book apparently guided by the instructions in the book. Considering the finding that the coursebooks seemed to have the potential to influence the teachers' pedagogical beliefs, this could be considered as a positive effect of the system.

Overall, the findings of this study support the general consensus in the literature that teacher-related factors such as their beliefs and their perceptions of tests play an

important role in the generation of washback (Alderson & Wall, 1993; Alderson & Hamp-Lyons, 1996; Badger & Yan, 2012; Burrows, 2004; Shih, 2009; Wall & Alderson, 1993; Wall & Horak, 2011; Wang, 2010; Watanabe, 2004a, Watanabe, 2004b). Bearing in mind the crucial role of the given contextual factors, however, it is clear that there is also an interplay between teacher-related factors and contextual factors. Thus, it appears that, as has been pointed out by other researchers, it is not possible to understand the mechanism of washback without taking the whole context into consideration (Cheng, 2004; Watanabe, 2004a).

5.3. Implications of the Study

The results of this study indicated that washback on teaching methodology occurred as a result of a complex interplay between different factors such as teaching materials, teacher-related factors and institution-specific factors. Thus, in line with the general consensus in the literature, the major implication of this study is that it is crucial to take account of both contextual and teacher-related factors in order to gain a deeper understanding of the mechanism of washback (Alderson & Wall, 1993; Cheng, 2004; Wang, 2010; Watanabe, 2004a). Specifically, with regard to the teachers' perceptions of the tests, the present study found that within the given context, the main driver of washback was the teachers' perceptions of the tests with regard to their consequences. Thus, the implication of this finding is that it might be beneficial for washback researchers to take into account the test participants' perceptions of the tests regarding their consequences before deciding that a given test is low-stakes and unlikely to induce washback, which might result in greater understanding of the washback effects of classroom assessment tests.

The most significant implications of the study, however, concern decision-makers at the given institution and coursebook designers. Considering that the institutional efforts to standardize teaching and control teacher behavior through the teaching materials and exams had both positive and negative effects on teaching, it appears that it might prove highly beneficial to find a middle ground between strict control of teacher behavior and allowing each teacher full rein, which could be achieved simply by involving the teachers in the decision-making processes including the designing of the pacing schedules and the tests. Also, incorporating alternative assessment tools into the system might help personalize the teaching and learning

process. Such changes to the system could help ameliorate the negative influence of the system on teaching practices while enhancing its positive effects.

With respect to the implications for coursebook designers, given the potential of coursebooks to influence teachers' beliefs and their methodologies, coursebook designers could help shape teachers' beliefs and lead their teaching in the right direction by designing interesting books that accord with the latest trends in ELT, and they could leverage this potential by designing functional, teacher-friendly teacher books that provide teachers with concrete guidance as to how to teach those coursebooks. Also, providing ongoing, face-to-face support and training would be highly beneficial.

5.4. Suggestions for Future Research

As stated before, the present study did not involve analyses of the tests themselves but relied on the participants' perceptions of the styles, properties and assessment objectives of the tests. It might be beneficial for future researchers to incorporate an analysis of test properties into their studies, which might provide a more objective and comprehensive understanding of the role of teachers' perceptions of a given test and the potential interaction of such perceptions with their pedagogical beliefs in the generation or inhibition of washback.

Also, the present study adopted a qualitative multiple case study approach, which does not allow making generalizations. Thus, further research could be conducted using quantitative methods in addition to qualitative methods and across a larger population for generalization purposes. Also, comparative studies that compare teachers' perceptions with those of their students could help understand the student's role in the generation of washback, which in turn, could provide a more comprehensive understanding of the role of contextual factors and thus the mechanism of washback in a given context.

Finally, one of the most significant findings that emerged from this study was the potential of coursebooks to influence teachers' pedagogical beliefs. Considering that teachers' beliefs play a crucial role in shaping their pedagogical decisions, further research that places particular focus on the role of coursebooks in the development of teachers' pedagogical beliefs would help advance our understanding of both the nature of teachers' pedagogical beliefs and the mechanism of washback.

6. REFERENCES

Alderson, J. C. (2004). The shape of things to come: Will it be normal distribution? In M. Milanovic & C.J. Weir (Eds), *European language testing in a global context: Proceedings of the ALTE Barcelona conference, July 2001*, (Studies in language testing 18) (pp. 1-26). Cambridge: Cambridge University Press.

Alderson, J. C., & Banerjee J. (2001). Language testing and assessment: State of the art: Part one. *Language Teaching, 34*(4), 213-236.

Alderson, J. C., Clapham C. M., & Wall , D. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Alderson, J.C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A washback study. *Language Testing*, *(13)*3, 280-297.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics. 14(*2), 115-129.

Anthony, E. M. (1963). Approach, method and technique. *Enlish Language Teaching. (17)*2, 63-67.

Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to research in education* (8th ed). Wadsworth: Cencage Learning.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Badger, R., & Yan, X. (2012). To what extent is communicative language teaching a feature of IELTS classes in China? *IELTS Research Reports Online Series*, *13*, 1-44. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_rr_volume13_report4.ashx

Bahat, S., Kemp, J., Shynin, E., Kara, S., Natur, N., Reis, S., et al. (1997). *Case study shel ehad mibatei haseifer shmishtatfim beproyekt, bagrut 2000* [*Alternative assessment: Case study of one school participating in Bagrut 2000*]. Seminar paper in research on language testing, Tel Aviv University.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing 13*(3), 257-279.

Bannister, D. (2003a). Kelly versus clockwork psychology. In F. Fransella (Ed.), *International handbook of personal construct psychology* (pp. 33-40). Chichester, UK: John Wiley & Sons.

Bannister, D. (2003b). The logic of passion. In F. Fransella (Ed.), *International handbook of personal construct psychology* (pp. 61-74). Chichester, UK: John Wiley & Sons

Berry, R. (2008). *Assessment for learning*. Hong Kong: University Press.

Black, P. (1998a). *Testing, Friend or Foe? The theory and practice of testing and assessment*. London: Falmer Press.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *The Phi Delta Kappan, 80*(2), 139-148. Retrieved from http://www.jstor.org/stable/20439383

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research context and methods* (pp. 113-128). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Chappell, P., Bodis, A., & Jackson, H. (2015). The impact of teacher cognition and classroom practices on IELTS test preparation courses in the Australian ELICOS sector. *IELTS Research Report Series, 6*, 1-61. Retrieved from http://www.ielts.org/researchers/research.aspx

Chappuis, J. 2009. *Seven strategies of assessment for learning*. Portland, OR: ETS Assessment Training Institute.

Cheng, L. (1995). *How does washback influence teaching? Implications for Hong Kong*. ERIC. Retrieved from ERIC database. (ED385143)

Cheng, L. (1999). Changing assessment: Washback on teacher perspectives and actions. *Teaching and Teacher Education, 15*(3), 253-271.

Cheng, L. (2004). The washbackback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research Contexts and Method* (pp. 147-170).

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-17). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Cheng, L. Curtis, A. (2012). Test impact and washback: Implications for teaching and learning. In C. Coombe, P., Davidson, B. O' Sullivan and S.Stoyynoff (Eds), *The Cambridge guide to second language assessment* (pp. 89-95). Cambridge: Cambridge University Press.

Chiari, G., & Nuzzo, M. L. (2003). Kelly's philosophy of constructive alternativism. In F. Fransella (Ed.), *International handbook of personal construct psychology* (pp. 41-50). Chichester, UK: John Wiley & Sons.

Chiari, G., & Nuzzo, M. L. (2010). *Constructivist psychotherapy: A narrative hermeneutic approach.* London: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education.* London: Routledge/ Falmer**.**

Coombe, C., Folse, K., & Hubley, N. (2007). *A practical guide to assessing English language learners.* Ann Arbor, MI: University of Michigan Press.

Çakıldere, B. (2013). *Washback effects of high-stakes exams, KPDS and ÜDS (YDS), on language learning of academic personnel (Nevşehir case study)* (Master's thesis). Gazi University, Ankara.

Dağtan, E. (2012). *An examination of the washback effect of the Interuniversity Board Foreign Language Test (The Example of Dicle Üniversity)* (Master's thesis). Dicle University, Diyarbakır.

Department of Education and Science (1988). Task group on assessment and testing-national curriculum: A report. London: Author.

Duran, Ö. (2011). *Teachers' and students' perceptions about classroom-based speaking tests and their washback* (Master's thesis). Bilkent University, Ankara.

Farrell, T. S. C. (2015). *Reflective language teaching: From research to practice.* London, New York: Bloomsbury.

Ferman, I. (1998). The impact of a new English foreign language oral matriculation test on the educational system (Master's thesis). Tel Aviv University.

Fransella, F., & Neimeyer, R. A. (2003). George Alexander Kelly: The man and his theory. In F. Fransella (Ed.), *International handbook of personal construct psychology* (pp. 21-32). Chichester, UK: John Wiley & Sons.

Frederiksen, J. R., & Collins, A. (1989). A Systems Approach to Educational Testing. *Educational Researcher, 18*(9), 27-32.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment.* London and New York: Routledge.

Glover, P. (2006). Examination influence on how teachers teach: A study of teacher talk (Doctoral dissertation). Department of Linguistics and Modern English Language, Lancaster University, Lancaster, England.

Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly, 3*(4), 333–368.

Guba, E.G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal 29* (1981), 75–91.

Hall, G. (2011). *Exploring English language teaching: Language in action.* New York: Routledge.

Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics 15*, 212-226.

Hamp-Lyons, L. (1997). Washback, impact and validity: ethical concerns. *Language Testing (14)*3, 295-303.

Hancock, C.R. (1994). *Alternative assessment and second language study: What and why?* ERIC. Retrieved from ERIC database. (ED 376695).

Heaton, J. B. (1990). *Classroom Testing.* Harlow: Longman.

Hughes, A. (1993). *Backwash and TOEFL 2000.* Reading, UK: University of Reading.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Kelly, G. A. (1955). *The psychology of personal constructs*. New York: Norton.

Kelly, G. A. (2003). A Brief introduction to personal construct theory. In F. Fransella (Ed.), *International handbook of personal construct psychology* (pp. 3-20). Chichester, UK: John Wiley & Sons

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), 116-134. Retrieved from http://journals.umcs.pl/lsmll/article/view/2167/2877

Köktürk, S. (2015). *Washback effects of examinations on teaching and learning English in Gazi University Preparatory School* (Master's thesis). Gazi University, Ankara.

Kumaravadivelu, B. (2006). *Understanding language teaching: From method to post-method.* Mahwah, NJ: Routledge.

Lankshear, C., & Knobel, M. (2004). *Handbook for teacher research: From design to implementation.* Maidenhead: Open University Press.

Larsen-Freeman, D. 2000. *Techniques and Principles in Language Teaching* (2nd edition) Oxford: Oxford University Press.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute-by-minute and day-by-day. *Educational Leadership, 63(3),* 18-24.

Madaus, G.F. (1988). The Distortion of Teaching and Testing: High-Stakes Testing and Instruction. *Peabody Journal of Education (65)*3, 29-46.

McNamara, T. (2000). *Language testing.* Oxford, UK: Oxford University Press.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, California: Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (Ed), *Educational measurement* (3rd edition, pp. 13-103). New York: Macmillan.

Messick, S. (1996). *Validity and washback in language testing*. (Report No. ETS-RR-96-17) Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 403 277)

Munby, H. (1981). *The place of teachers' beliefs in research on teacher thinking and decision making, and on alternative methodology* (Report No. 9042) Austin, TX: Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED214844).

Munoz, A. P., & Alvarez, M.E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing, 27*(1), 33-49.

Olovsson, T.G. (2015) The assessment process in a Swedish year five classroom: 'Reach page 52!', *Education (3)*13, 743-755 DOI: 10.1080/03004279.2014.899382

Özmen, K. S. (2011). Washback effect of the Inter-university Foreign Language Examination on foreign language competences of candidate academics. *Novitas-ROYAL (Research on Youth and Language), 5*(2), 215-228.

Patton, M.Q. (1999). Enhancing the quality and credibility of qualitative analysis. *HSR: Health Services Research (34)*5 Part II, 1189-1202.

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, *68*, 679-682.

Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria: Association for Supervision and Curriculum Development.

Proficient. In *Oxford Avanced Learner's Dictionary* (2018). Retrieved from https://www.oxfordlearnersdictionaries.com/definition/american_english/proficient#proficient__29

Qi, L. (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: Principles, Policy & Practice (14)*1, 51-74

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge: Cambridge University Press.

Richards, J. C., & Rogers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed). Cambridge: Cambridge University Press.

Rowley, J. (2002). Using case studies in research. *Management Reseach News (25)*1, 16-27.

Sağlam, A. L. G. (2016). *Washback and instructional sensitivity of a theme-based high-stakes English language proficiency test* (Master's thesis). Yeditepe University, İstanbul.

Scott, John. 2000. Authentic assessment tools. *Using Authentic Assessment in Vocational Education Information Series 381*, 33-48.

Sevimli, S. E. (2007). *The Washback effects of foreign language component of the University Entrance Examination on the teaching and learning context of English language groups in secondary education (a case study)* (Master's thesis). University of Gaziantep, Gaziantep.

Shaw, M. L. G., & Thomas, L.F. (1978). FOCUS on education--an interactive computer system for the development and analysis of repertory grids. *Int. J. Man-Machine Studies10*, 139-173

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information 22*, 63–75

Shepard, L.A. (1994). The challenges of assessing young children appropriately. *Phi Delta Kappan, 76*, 206-212.

Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning*. (Report No. CSE-TR-517). Los Angeles, CA: CRESST. (ERIC Document Reproduction Service No. ED443880).

Shih, C. M. (2007). A new washback model of students' learning. *Canadian Modern Language Review, 64*(1), 135–162.

Shohamy, E. 1993. The power of tests: the impact of language tests on teaching and learning. *NFLC Occasional Papers*, Washington DC: National Foreign Language Center.

Shohamy, E., Donitza-Schmidt, S.,& Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing (13)*3, 298-317.

Shohamy, E. 1998. Inside the "black box" of classroom language tests. *Studies Anglica Posnaniensia 33*, 343-352. (special issue in honor of Kari Sajaavara).

Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education (14)*1, 117-130.

Smith, M. A. (1980). The design, analysis and interpretation of repertory grids. *Int. J. Man-Machine Studies 13*, 3-24.

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.

Stiggins, R. J, & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.

Stiggins, R. J. (1992). High quality classroom assessment: What does it really mean? *Educational Measurement: Issues and Practice, 11*(2), 35–39.

Stiggins, R. J. (2000). Classroom assessment: *A history of neglect, a future of immense potential*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Şentürk, F. (2013). *Washback effect of KET exam in learning English as a foreign langauge* (Master's thesis). Çağ University, Mersin.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics 29*, 21–36.

Thornbury, S. (2006). *An A-Z of ELT*. Oxford: MacMillan.

Wall, D. (1997). Test impact and washback. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language education: Vol. 7. Language testing and evaluation* (pp. 291–302). Dordrecht, the Netherlands: Kluwer.

Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10*, 41–69

Wall, D., & Horak T. (2006). The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 1, the baseline study. (MS 34) Princeton, NJ: Educational Testing Service

Wall, D., & Horak, T. (2008). The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 2, Coping with change (TOEFL IBT-05). Princeton, NJ: Educational Testing Service.

Wall, D. & Horak, T. (2011). The impact of changes in the TOEFL exam on teaching in a sample of countries in Europe and learning in Central and Eastern Europe. Phase 3, The role of the coursebook. Phase 4, Describing change. (TOEFL IBT-17). Princeton, NJ: Educational Testing Service.

Wang, J. (2010). *A study of the role of the 'teacher factor' in washback* (Doctoral dissertation). Retrieved from ProQuest (NR74872).

Watanabe, Y. (2004a). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis, (Eds.), *Washback in Language Testing: Research Context and Methods* (pp. 19-36). Mahwah, NJ: Lawrence Erlbaum Associates.

Watanabe, Y. (2004b). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis, (Eds.), *Washback in Language Testing: Research Context and Methods* (pp. 129-146). Mahwah, NJ: Lawrence Erlbaum Associates.

Yıldırım, Ö. (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. 12*(2), 92-116. Retrieved from http://www.asianefl-journal.com

Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In B. Wildemuth (Ed.), *Applications of Social Research Methods to Questions in Information and Library Science* (pp. 308-319). Westport, CT: Libraries Unlimited.

7. APPENDICES

Appendix 1: Participant Consent Form

Dear Colleague,

You are invited to participate in a research study entitled "**The impact of institutional tests on teachers' teaching methodologies,**" which I am conducting for my Master's thesis.

Purpose of the Study

The overall purpose of the study is to explore whether and how institutional tests influence teachers' teaching methodologies.

Data Collection Instruments

Data will be collected from you through the following instruments:

1. The repertory grid technique
2. Classroom observations of your lessons for 4 teaching hours
3. Semi-structured post-observation interviews

Also, the classroom observations and the post-observation interviews will be audio-recorded.

Confidentiality

Your identity will only be known to the researcher. The results of this study will be used for a Master's thesis and may be used in conference presentations or reports; however, your name will not be revealed under any circumstances. Also, only the researcher will have access to the data you provide, and you may choose to withdraw from the study at any time, in which case all the data you have provided will be deleted. Finally, you will be given a copy of this signed consent form.

Participant's Consent

I have read all the information above. All my questions have been answered to my satisfaction. I voluntarily agree to participate in this study.

Participant's name:

Signature:

Contact email:

Date:

Appendix 2: A Sample Repertory Grid

Construct №	Triads	Emergent Constructs (Similarities)	Rating Scale 1 2 3 4 5									Implicit Constructs (Contrasts)
			E1	E2	E3	A1	A2	A3	I1	I2	I3	
1	(E1) (E2) I3	Presents the content in a well-organized way.	1	2	2	3	4	3	4	4	5	Jumps from one idea to another; totally disorganized.
2	(I3) (I1) E2	Lectures throughout the lesson.	5	5	4	4	3	3	1	2	1	Encourages a lot of discussion.
3												

Appendix 3: Semi-structured Interview Guide

1. What do you think the most important part of learning English is (grammar, vocabulary, listening, reading etc)?

2. What aspects of the English language do you focus on the most in your daily practice? Why? How do you teach them?

3. Considering the course you teach, what do you think the exams administered at the Preparatory School aim to assess?

4. What do you think the scores your students get on the exams show?

5. How would you like to teach the language if you had complete freedom as a teacher?

Appendix 4: Permission Letter

<div align="center">

ÇAĞ ÜNİVERSİTESİ

HAZIRLIK OKULU MÜDÜRLÜĞÜ'NE

</div>

MERSİN

05.02.2018

Çağ Üniversitesi Sosyal Bilimler Enstitüsü İngiliz Dili Anabilim Dalı yüksek lisans çalışmam kapsamında, "Kurumsal sınavların öğretmenlerin öğretim metodolojileri üzerine etkisi" konulu araştırmam için kurumunuzda görevli üç öğretim görevlisinin sınıflarında, her biri için dörder saat olmak üzere, 19 Şubat-16 Mart 2018 tarihleri arasında sınıf gözlemi yapmak istiyorum.

Bilgilerinize ve onaylarınıza arz ederim.

Songül Yeliz SARI

— Uygundur

Hamdi ÖNAL
Çağ Üniv. Hazırlık Okulu
Müdürü

— İslme e gürüş

—Doç. Dr. Murat KOÇ
Sos. Bil. Enst. Müd.

Appendix 5: Ethics Committee Permission Document

T.C
ÇAĞ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
TEZ / ARAŞTIRMA / ANKET / ÇALIŞMA İZNİ / ETİK KURULU İZİNİ TALEP FORMU VE ONAY TUTANAK FORMU

T.C. NOSU	26957336832
ADI VE SOYADI	Songül Yeliz SARI
ÖĞRENCİ NO	20118016
TEL. NO.' LARI	0533 541 07 57
E - MAİL ADRESLERİ	syelizsari@gmail.com
ANA BİLİM DALI	İngiliz Dili Eğitimi
PROGRAM ADI	İngiliz Dili Eğitimi Yüksek Lisans Programı
BİLİM DALININ ADI	
HANGİ AŞAMADA OLDUĞU (DERS / TEZ)	Tez
İSTEKDE BULUNDUĞU DÖNEME AİT DÖNEMLİK KAYDININ YAPILIP-YAPILMADIĞI	2017-2018 BAHAR......................... dönemi kaydımı yeniledim.

TEZİN KONUSU	Kurumsal sınavların öğretmenlerin öğretim metodolojileri üzerine etkisi
TEZİN AMACI	Kurumsal sınavların öğretmenlerin öğretim metodolojileri üzerinde ne tür etkileri olduğunu incelemek
TEZİN TÜRKÇE ÖZETİ	Bu çalışmanın iki amacı vardır. Birinci amacı, kurumsal gelişim sınavlarının öğretmenlerin öğretim metodolojileri üzerine geriye dönük etkisini incelemektir. Çalışmanın ikincil amacı ise öğretmenlerin pedagojik inançlarının ve sınavlarla ilgili algılarının sınavların geriye dönük etkisinin ortaya çıkmasındaki rolünü incelemektir. Bağlamsal etkenlerin sınavların geriye dönük etkisinin ortaya çıkmasındaki kritik rolü göz önünde bulundurularak, çalışma niteliksel çoklu vaka çalışması şeklinde tasarlanmıştır. Çalışma Türkiye'de bir üniversitenin İngilizce hazırlık okulunda yapılmıştır. Çalışmanın katılımcıları söz konusu kurumda çalışan 3 öğretim görevlisinden oluşmaktadır. Veriler, birbiriyle ilintili üç aşamada, repertory grid tekniği, sınıfiçi gözlemleri ve yarı-yapılandırılmış derinlemesine görüşme yöntemiyle toplanmıştır. Repertory grid verilerinin analizi cluster analizi yöntemiyle; sınıfiçi gözlemleri ve görüşmeler aracılığıyla toplanan veriler ise, içerik çözümlemesi tekniğiyle analiz edilmiştir. Sonuçlar, sınavların öğretim içeriğinin yanı sıra, öğretmenlerin öğretim metodolojileri üzerine de geriye dönük etkileri olduğunu göstermiştir. Ancak, bulgular, bu etkilerin ana sebebinin sınavlardan ziyade, öğretmenlerin sınavlarla ilgili algıları, bağlamsal etkenler ve öğretmenlerin pedagojik inançları arasındaki karmaşık bir etkileşim sonucu ortaya çıktığına işaret etmektedir.
ARAŞTIRMA YAPILACAK OLAN SEKTÖRLER / KURUMLARIN ADLARI	Çağ Üniversitesi
İZİN ALINACAK OLAN KURUMA AİT BİLGİLER (KURUMUN ADI - ŞUBESİ / MÜDÜRLÜĞÜ - İLÇESİ)	İL) Çağ Üniversitesi/Hazırlık Okulu/ Yenice/ Mersin
YAPILMAK İSTENEN ÇALIŞMANIN İZİN ALINMAK İSTENEN KURUMUN HANGİ İLÇELERİNE/ HANGİ KURUMUNA/ HANGİ BÖLÜMÜNDE/ HANGİ ALANDA/ HANGİ KONULARDA/ HANGİ GRUBA/ KİMLERE/ NE UYGULANACAĞI GİBİ AYRINTILI BİLGİLER	Çağ Üniversitesi Hazırlık Okulunda görevli üç öğretim görevlisiyle vaka analizi
UYGULANACAK OLAN ÇALIŞMAYA AİT ANKETLERİN/ ÖLÇEKLERİN BAŞLIKLARI/ HANGİ ANKETLERİN - ÖLÇELERİN UYGULANACAĞI	Repertory Grid/ Sınıf gözlemleri/ Yarı yapılandırılmış görüşme
EKLER (ANKETLER, ÖLÇEKLER, FORMLAR .... GİBİ EVRAKLARIN İSİMLERİYLE BİRLİKTE KAÇ ADET/SAYFA OLDUKLARINA AİT BİLGİLER İLE AYRINTILI YAZILACAKTIR)	1) Repertory Grid (3 adet) 2) 5 sorudan oluşan yarı yapılandırılmış görüşme (3 adet) 3) Katılımcı gönüllülük formu (3 adet)

ÖĞRENCİNİN ADI - SOYADI: SARI Songül Yeliz ÖĞRENCİNİN İMZASI: .... TARİH: 31./10./20.18

1. Seçilen konu Bilim ve İş Dünyasına katkı sağlayabilecektir.

2. Anılan konu .... faaliyet alanı içerisine girmektedir.

1.TEZ DANIŞMANININ ONAYI
Adı - Soyadı: Prof. Dr. ....
Unvanı :
İmzası :
....../......./20....

2.TEZ DANIŞMANININ ONAYI (VARSA)
Adı - Soyadı : ....
Unvanı :
İmzası :
....../......./20....

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRÜNÜN ONAYI
Adı - Soyadı: ....
Unvanı :
İmzası :
......7./20....

A.B.D.BAŞKANININ ONAYI
Adı - Soyadı: ....
Unvanı :
İmzası :
....../......./20....

ETİK KURULU ASIL ÜYELERİNE AİT BİLGİLER

Adı - Soyadı: Mustafa BAŞARAN	Adı - Soyadı: Yücel ERTEKİN	Adı - Soyadı: Deniz Aynur GÜLER	Adı - Soyadı: Ali Engin OBA	Adı - Soyadı: Mustafa Tevfik ODMAN	Adı - Soyadı: ....
Unvanı : Prof. Dr.	Unvanı : Prof. Dr.	Unvanı : Prof. Dr.	Unvanı : Prof. Dr.	Unvanı : Prof. Dr.	Unvanı : ....
İmzası :	İmzası :	İmzası :	İmzası :	İmzası :	İmzası : ....
....../......./20....	....../......./20....	31./10./2018	....../......./20....	....../......./20....	....../......./20....
Etik Kurulu Jüri Başkanı - Asıl Üye	Etik Kurulu Jüri Asıl Üyesi	Etik Kurulu Jüri Asıl Üyesi	Etik Kurulu Jüri Asıl Üyesi	Etik Kurulu Jüri Asıl Üyesi	Etik Kurulu Jüri Yedek Üyesi

Adı - Soyadı: ....

Unvanı :

İmzası :
....../......./20....

Etik Kurulu Jüri Yedek Üyesi

OY BİRLİĞİ İLE ◯

OY ÇOKLUĞU İLE ⊘

Çalışma yapılacak olan tez için uygulayacak olduğu Anketleri/ Formları/Ölçekleri Çağ Üniversitesi Etik Kurulu Asıl Jüri Üyelerince incelenmiş olup, .../.../20.. - .../.../20.. tarihleri arasında uygulanmak üzere gerekli iznin verilmesi tarafımızca uygundur.

AÇIKLAMA: BU FORM ÖĞRENCİLER TARAFINDAN HAZIRLANDIKTAN SONRA ENSTİTÜ MÜDÜRÜNE ONAYLATILARAK ENSTİTÜ SEKRETERLİĞİNE TESLİM EDİLECEKTİR.

EKLERİ: .... (............) Sayfa .... Ölçeği.
.... (............) Sayfa .... Anket.
.... (............) Sayfa .... Formları.

SoBE— 1732

8. CURRICULUM VITAE

Songül Yeliz SARI was born in 1983. She earned her Bachelor of Arts degree in English Language and Literature from Hacettepe University in 2004. She has been working as an English instructor at Çağ University since 2004.

Contact Email Address: syelizsari@gmail.com