Washback and Instructional Sensitivity of a Theme-Based High-Stakes English Language

Proficiency Test

Dissertation Submitted to the Institute of Educational Sciences

In Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy
in
English Language Education

by

Aslı Lidice Gokturk Saglam

Yeditepe University

2016

YEDİTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ TESLİM ve ONAY TUTANAĞI

KONU: Washback and Instructional Sensitivity of a Theme-Based High-Stakes English Language Proficiency Test
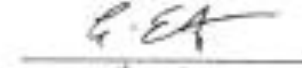
ONAY:

Doç. Dr Hassan Farhady
(Danışman)

_____
(İmza)
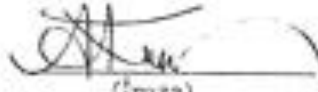
Prof. Dr Ayşe Semra Akyel
(Üye)

_____
(İmza)
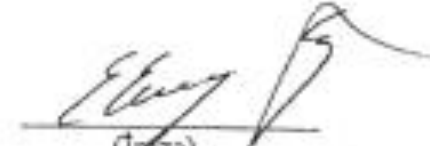
Doç. Dr. Naciye Gülcan Ercetin Koca
(Üye)

_____
(İmza)

Yard. Doç. Dr. Orhan Kocaman
(Üye)

_____
(İmza)

Yard. Doç. Dr Evrim Eveyik Aydın
(Üye)

_____
(İmza)

TESLİM EDEN   : Aslı Lidice Gültürk Sağlam
TEZ SAVUNMA TARİHİ : 02.06.2016
TEZ ONAY TARİHİ   : 02.06.2016

ACKNOWLEDGMENTS

My deepest appreciation goes to all of the teachers who were so open to sharing their ideas and teaching practice during this study.

I would also like to thank my parents, my brother and sister as well as my selected family-friends for their constant love and support through this process.

I would like to acknowledge my husband Zafer and kids, Mirza and Ela, in thanks for being understanding when I was too busy. Without their words of encouragement, I would not have been able to finish what I started.

Thank you all.

VITA
ASLI LİDİCE GÖKTÜRK SAGLAM

EDUCATION

2010 - 2015	Ph.D., Yeditepe University, Department of English Language Education, Istanbul, Turkey.
1998 - 2002	M.A., Middle East Technical University, Department of English Language Teaching, Ankara, Turkey
1994 - 1998	B.A., Middle East Technical University, Department of English Language Teaching, Ankara, Turkey

WORK EXPERIENCE

2008- Present	Instructor, Ozyegin University Istanbul, School of Languages Preparatory English Program, Istanbul, Turkey.
2006-2007	Instructor, Curriculum Advisor, Bahcesehir University, Language Preparatory School, Istanbul, Turkey.
2003-2006	Instructor, Curriculum Advisor, Bilgi University, Foundation Year English Program, Istanbul, Turkey.
2003-2004	Instructor, Coordinator, Bilgi University, Social Sciences Institute English Preparatory Program, Istanbul, Turkey.
2000-2002	Instructor, Level Coordinator, Bilgi University, Language Preparatory School, Istanbul, Turkey.
1998 –2000	Instructor, Bilkent University, Language Preparatory School, Ankara, Turkey.

PUBLICATIONS

Saglam, A. & Sert, S. (2012). Perceptions of In-Service Teachers Regarding Technology Integrated English Language Teaching, Turkish Online Journal of Qualitative Inquiry, 3(3), 1-14.

Saglam, A. (2012). Using Students' Evaluations to Measure Educational Quality". Procedia - Social and Behavioral Sciences, 70 (25), 1009–1015.

Saglam, A. & Sirin, P. (2012). Unveiling the Mystery: Investigating L2 Reading Motivation in an EFL Classroom. Proceedings of 11th METU Convention.

CONFERENCE PRESENTATIONS

Saglam, A. (2015). A study of the washback effects of a Theme-Based Language Proficiency Test on teaching and learning English at tertiary level in Turkey, Paper presented at 12th EALTA Conference, University of Copenhagen, Copenhagen, Denmark

Saglam, A. (2015). Blogging an Adventure in Professional Development, Paper presented at 49th Annual International IATEFL Conference and Exhibition, Manchester, UK.

Saglam, A. (2012). Exploring Tertiary-Level English Teachers' Practical Knowledge, Paper presented at 7th International ELT Research Conference, Onsekiz Mart University, Çanakkale, Turkey.

Saglam, A. (2012). Using Students' Evaluations to Measure Educational Quality, Paper presented at Akdeniz Language Studies Conference 2012, Antalya, Turkey

Saglam, A. (2010). Renegotiating Standards in Speaking Through Peer and Self-Assessment, Paper presented at IATEFL Testing SIG Event, Zayed University, Dubai, UAE.

Saglam, A. (2010). Extending Borders of Classroom: Using Podcasts to Optimize Language Learning, Paper presented at 31st Annual TESOL Greece International Convention, Hellenic American Union, Athens, Greece.

Saglam, A. (2010). Functionalising Diagnostic Tests in Curriculum and Assessment: A Cognitive Approach, Paper presented at IATEFL Slovenia, Ljubljana, Slovenia.

PROFESSIONAL ACTIVITIES

Member of European Association for Language Testing and Assessment, 2008- Present

Member of INGED, 1998- Present

RESEARCH AND TEACHING INTERESTS

Testing and Assessment, Technology in Education, Teacher Education

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Washback and Instruction Sensitivity of a Theme-Based High-Stakes English Language

Proficiency Test

By

Aslı Lidice Göktürk Sağlam

The purpose of this study is to examine the potential washback effect, instructional sensitivity and consequential validity of a local theme-based high-stakes English language proficiency test that is implemented in tertiary education in a Turkish context and reach a better understanding of construct of integrated language assessment tasks in an EAP context. The study aimed at exploring the validity of the assumption that employing an integrated theme-based test of English language proficiency, which is similar to authentic language use in the tertiary education context in Turkey, would bring about a positive test influence upon learning. The study comprised of two parts and for the first part of the research study which examined washback and instructional sensitivity of TRACE, 21 language instructors teaching at the upper-intermediate and advanced level in the Preparatory English Program and 147 EFL students participated in the study. As for the second part of the study which investigated viability of test decisions over time, 39 freshman students in mainstream courses and 19 university instructors who offered courses in the departments participated in the study. A mixed-method research design was adopted. Thus, both qualitative and quantitative data were

obtained through questionnaires, semi-structured interviews related to the perceptions of the students and the teachers, classroom observations and the pre- and post-proficiency exam scores of the preparatory students. The study mainly found that the test exerted positive and negative effects on teaching and learning. The findings of the study revealed significant implications with respect to the ELT teachers, teacher trainers and test designers.

Key words: washback, instructional sensitivity, consequential validity, theme-based language proficiency test, integrated language proficiency test

KISA ÖZET

Bir Temaya Odaklanan Bütünleştirilmiş Bir Ingilizce Dil Sınavının Geriye Dönük Etkisi

Aslı Lidice Göktürk Sağlam

Bu çalışmanın amacı Türkiye'de üniversite seviyesinde uygulanan yerel ve bir temaya odaklanan bütünleştirilmiş bir Ingilizce Dil sınavının potansiyel geriye dönük etkisini, öğretimsel etkisini ve sonuç geçerliğini incelemektir. Araştırma sonucunun akademik amaçlı İngilizce bağlamında bütünleştirilmiş dil sınavları uygulanması için daha iyi bir kavramsal anlayış sağlayacağı umulmaktadır. Bu çalışma bütünleştirilmiş dil sınavlarının öğrencilerden gerçek öğrenim hayatlarında onlardan beklenilen akademik becerileri yansıttığı için onların başarılarına olumlu geri dönük etkisi olabileceği varsayımını araştırmayı amaçlamaktadır.

geliştirmek ve değerlendirmektir. Çalışma, geriye dönük etki ve sonuç geçerliliği olmak üzere iki bölümden oluşmaktadır. Çalışmanın geriye dönük sınav etkileri ve öğretimsel etkileri araştıran ilk bölümünde, veriler katılımcıların bu konudaki algıları ile ilgili nicel veriler anket, yarı yapılandırılmış röportaj ve sınıf gözlemi yöntemiyle elde edilmiştir. Nitel veriler hazırlık sınıfı öğrencilerinin dil yeterlilik sınavı ön-test ve son-test sonuçlarından elde edilmiştir. 21 Hazırlık Programı dil eğitmeni ve 147 ortanın üstü ve ileri düzey Hazırlık Programı öğrencisi katılmıştır. Sonuç geçerliliğini araştıran ikinci bölümde ise katılımcılardan yarı yapılandırılmış röportaj ve anket yolu ile bilgi toplanmıştır. Bu bölüm için hazırlık programını bitirip bölüme geçen 39 birinci sınıf öğrencisinden ve onların bölümde ders aldıkları 19 universite hocasından bilgi toplanmıştır. Çalışma dil sınavının olumlu ve olumsuz geriye dönük etkileri olduğunu ortaya çıkarmıştır. Çalışmanın bulguları dil Hazırlık Programı geliştirme ve değerlendirmeye ve öğretmen eğitimine yönelik önemli sonuçlar ortaya koymuştur.

Anahtar Kelimeler: geriye dönük etki, öğretimsel etki, sonuç geçerliği, bütünleştirilmiş dil sınavı, tema-odaklı dil sınavı

CHAPTER 1

INTRODUCTION

1.1 . Background to the Study

Second language assessment has traditionally focused on measuring four skills: speaking, listening, reading, and writing as interdependent constructs. However, such a focus may fall short in the depiction of real-world communicative acts displaying language proficiency, since these rely on integration of two or more of these skills as well as non-linguistic cognitive skills. Inclusion of integrated tasks of listening, reading, speaking and writing in a proficiency test can better cater for the demands of authentic communication and reflect the test-takers performance. For Cumming (2013) the construct of integrated assessment mirror academic literacy activities and therefore the assessment tasks are more authentic and sufficiently challenging. One gate keeper which determines language achievement of the students for a Turkish university is the Test of Readiness of Academic English (TRACE) in that its purpose is to differentiate between language learners who have the ability to use English for academic purposes in university classrooms and those who do not. TRACE is an institutional theme-based integrated language proficiency test which presupposes that the interactive nature of its listening, reading and writing tasks will be highly indicative of target language use (TLU) in academic classrooms. In other words, as an exam which assesses language proficiency via authentic theme-based integrated assessment tasks, it may be more valid since it mirrors the constructs of real-life demands of academic life. Thus, test decisions based on TRACE are assumed to lead to inferences about a test-taker's academic performance and their ability to engage in and sustain academic work (TRACE Specifications, 2008).

A theme based proficiency test which encompasses integrated tasks may also have a positive washback effect on both the instructional process and products in terms of the achievement of learning. This stems from the fact that the focus of instruction on synthesizing content of reading, listening, speaking and using this information in writing is a common assignment in the academic course work of English as a Second Language programs in many preparatory schools at the tertiary level.

The present research aims at investigating whether, as an integrated theme based proficiency test, TRACE is sensitive to instruction. Research studies to date into the use of topic-based tests of language proficiency (Fox, 2004; Jennings et.al.,1999; Farhady & Sabeti Dianati, 2000) and the use of integrated tasks as a measure of second language proficiency are limited in number, and the latter has examined predominantly integrated tasks as a measure of writing ability rather than language proficiency (Cumming et al., 2006; Plakan, 2009), reading ability (Cohen, 1998; Yu, 2007, 2008, 2009 in Frost, Elder & Wigglesworth, 2011), and speaking ability (Lee, 2006; Frost, Elder & Wigglesworth, 2011). To address this research gap, this study examines the effect of the instruction on TRACE, the washback of the test on instruction and participants, and the viability of test decisions over time after students meet the exit criteria. Applying a mixed method approach to investigate the instruction sensitivity of the TRACE, a variety of data collection instruments were used and data was analysed both qualitatively and quantitatively. Participants of the study entailed students of preparatory school who were placed in advanced and upper-intermediate levels, freshman students, English language teachers of the Preparatory English Language Program (PEP) and university professors of mainstream courses in the 2013-2014 academic year. In addition to interviews, questionnaires and classroom observations, the data collection involved pre- and post- proficiency test scores of these EFL students. A pre-test was administered at the beginning of a 4-month English language instruction period while a post-test was administered at the end of instruction.

1.   2. Statement of the Problem

Ozyegin University is a foundation university in Istanbul which uses English as the medium of instruction both at the preparatory and undergraduate programs. Prospective students are required to certify their English language competency before they start the undergraduate programs. The students are obliged to pass the TOEFL exam (with an average of 79), IELTS (with an average of 6.5) or the in-house written language proficiency exam of the PEP; TRACE (with an average of 65). Students who pass one of these exams continue their education at different departments. However, the ones who fail TRACE at the onset of an academic year are placed in advanced or upper-intermediate level with in the PEP on the basis of their level of proficiency and receive instruction in academic English. Based on their levels, students receive 20-16 hours of English instruction per week. The courses are designed to meet the students' academic language needs and they focus on a range of skills involving reading, writing, speaking, listening, grammar and vocabulary. Students of PEP come from various disciplines such as, Engineering, Law, Architecture, Culinary Arts, Aviation and Business Administration. The average number of students in a class is 16. At the end of instruction in PEP, students sit the institutional language proficiency exam, TRACE.

Tested skills and test format in TRACE models authentic language use since students are required to read multiple reading texts of different lengths and genre, respond to questions based on while and lecture listening and respond to a writing question by making use of their lecture notes about the very same theme exploited previously in reading and listening texts. TRACE is constructed upon the assumption that employing an integrated theme-based test of English language proficiency which is similar to authentic language use in the tertiary education context in Turkey would bring about positive test influence upon teaching and learning. Consequently, considering the discussions above, this study is an

attempt to examine instructional sensitivity and potential washback effect of as well as consequential validity of a theme based integrated language proficiency test and reach a better understanding of construct of integrated language assessment tasks in an EAP context.

1. 3. Purpose of the Study

This study aims at exploring the potential washback effect, instructional sensitivity and consequential validity of a theme based integrated language proficiency test and reach a better understanding of construct of integrated language assessment tasks in an EAP context. Previous washback studies suggested conclusions that are based on perceptions and recommendations of test-takers and educators. However, this study aimed at going one step beyond since it validates the claims of participants about washback of the test by examining the learning outcome. There were qualitative and quantitative measures to cross validate claims made and reach conclusions regarding how these claims are operationalized. Furthermore, examining viability of test decisions over time acted as a confirmatory study which followed up the potential washback of the TRACE and examined the consequential validity. This provided valuable information about how the test operates in the local context it is used in and shed light on how to make the best use of tests to engineer positive washback. It is commonly agreed that researching effects of a test may have connotations for educational administration, materials development, teacher training and resourcing as well as test development and revision (Wall and Alderson, 1993; Alderson and Hamp-Lyons, 1996; Shohamy et al., 1996; Watanabe, 1996). Therefore, Green (2007a, p.30) states: "It is important to gain ecologically grounded understandings of how a test operates within an educational context, rather than (or in addition to) seeking to isolate the effects of testing in experimental fashion". Pursuing 'ecologically grounded understandings' may result in comparisons between test data and instruction. As a result, this may reveal whether test scores increase and whether language and skills that are manifested in terms of objectives and described in the curriculum are acquired as a consequence of instructional practices. In this case, TRACE is

4

examined with respect to its potential to boost authenticity and validity as an integrated theme based test, reflecting real life demands of the academic context it operates in. Consequently, it had connotations with respect to materials development, teacher training and resourcing as well as test development and revision.

Furthermore, according to some testing researchers (Cumming, 2013; Yu, 2013) the construct of integrated assessment needs to be better defined due to several factors including difficulty of distinguishing; (1) the language produced by test-takers from source materials they are exposed in the test, and confounding effects of one language ability over another and (3) threshold level of linguistic and cognitive abilities that may affect performance of integrated tasks. Therefore, this study sets a useful agenda for inquiry and aims at reaching an understanding of integrated assessment by examining test characteristics and the quality of a theme based proficiency test for the tertiary level in assessing English for Academic Purposes in a Turkish context.

Specifically, the following research questions were addressed:

1. Is there a potential washback effect of the TRACE exam on teaching (methods, materials, and tasks)?

2. How does the language instruction program, based on EAP skills, result in gains in scores on the writing, listening, and reading parts of the TRACE exam?

3. How effective are the decisions made by TRACE in identifying the language competency required for academic study within this university program over the time when students embark on their further academic studies at their departments?

1. 4. Overview of Methodology

1. 4. 1. Participants

For the first part of the research study which examined washback and instructional sensitivity of TRACE, 21 language instructors teaching at the upper-intermediate and advanced

levels in the PEP and 147 EFL students who were enrolled in the program participated in the study.

As for the second part of the study which investigated viability of test decisions over time, data was gathered from 39 freshman students in mainstream courses and 19 university instructors of mainstream courses who offered courses in the departments.

1. 4. 2. Setting

The present study was conducted at the Preparatory English Language Program (PEP) at Ozyegin University in Istanbul, Turkey which is an English medium institution. The program is in pursuit of teaching English for general and academic purposes across different language competency levels in order to aid students to excel in their undergraduate academic studies. The overarching aim of the PEP is developing students with a good command of written and spoken English.

1. 4. 3. Data Collection Instruments

In an attempt to unveil the potential washback effect of the TRACE exam on teaching with regard to methods, materials, and tasks, questionnaires were used and interviews were conducted with both students and teachers of the PEP. A student questionnaire surveyed learner perceptions and a teacher questionnaire examined opinions of the proficiency exam, TRACE. Additionally, the information gathered was supported through semi-structured focus group meetings that were held with the students and one-to one interviews with the teachers. Finally, the data was screened against classroom observations.

As for the second research question, namely instructional sensitivity of TRACE, the pre- and post- proficiency scores of the students enrolled in the PEP were compared to identify if there was any difference in their language proficiency regarding 3 language skills (reading, listening and writing) after the implementation of the program. The gains in scores of the students who failed the first administration of the TRACE (pre- test) and were placed at upper-

intermediate and advanced levels within the program and gains in scores of the same students after receiving instruction were compared to find out whether any difference existed in their level of reading, writing, and listening proficiency. Furthermore, students' performance in pre- and post- test was compared to their perception of how much they think they learned in the program that was gathered through questionnaires and interviews.

Finally, to answer the last research question regarding consequential validity of TRACE and viability of these test-based decisions over time, semi-structured, one-on-one interviews were conducted with academics who offered departmental courses at the undergraduate level. In addition to identify teacher and student perceptions questionnaires were utilised.

1. 4. 4. Data Analysis

The present research study adopted a mixed method design approach in order to ensure internal validity. Data was gathered through a repertoire of a variety of data collection instruments including classroom observations, questionnaires, interviews and pre- and post- proficiency scores.

In an attempt to answer the first research question concerning potential washback effect of TRACE, student and teacher questionnaires were designed, piloted and analysed statistically to canvas the perceptions of the students and teachers of PEP.
In addition, the semi-structured interviews conducted with students and one-on-one interviews carried out with teachers were coded and transcribed adhering to Bogdan and Bikley's (1998) framework of pattern coding. Finally, classroom observations conducted to validate the data gathered were analysed both qualitatively and quantitatively.

As for the second research question, the pre- proficiency scores gained in the first administration of TRACE at the onset of the program and post proficiency scores gained in the second administration of the TRACE upon completion of the program were analysed statistically to find out whether TRACE was sensitive to instruction. Furthermore, the data were

supported via semi-structured student focus group interviews and one-on-one interviews with teachers given to the two groups of participants.

Furthermore, teacher and student questionnaires were tabulated and analysed statistically to investigate whether freshman students' and university instructors of undergraduate courses perceived TRACE-based decisions over time ensured academic success in EAP. The consequential validity of TRACE was also investigated through one-on-one interviews with university instructors who offered undergraduate courses and these interviews were also recorded, transcribed and coded with a focus on meaning coding, condensing the meaning and interpreting as discussed in detail in Bogdan and Biklen's model (1998).

1. 5. Contribution of the Study

It is often hypothesized that when there is a curricular alignment with in a language program between what is taught and what is tested washback is apt to be strong. TRACE functions in a local context in which such alignment is evident since TRACE consists of tasks that replicate target language use (TLU) domain. Consequently, it differs from other standardized high-stakes tests because it particularly aims at distinguishing test-takers who have the ability to use English for academic purposes in university classrooms. Therefore, investigating its washback is of vital importance for engineering positive washback and maintaining consequential validity.

Furthermore, this research study involved multiple participants who were surveyed through different data collection instruments and linked a wide array of variables together including test design, participants, teaching and learning processes and instructional sensitivity (product) as well as viability of test decisions to examine test effect through teaching and learning processes to proficiency test scores. This comprehensive approach to washback research has not been thoroughly emphasized in the current literature and this study aimed to fill this gap to investigate the complex nature of the phenomenon and trace evidence for consequential validity regarding the test-takers ability to engage in and sustain academic work.

1. 6. Limitations of the Study

Finally, the present study has limited external validity since it was carried out in a specific context, which prevents generalization of the results to different contexts.

Another limitation relates to one of the data collection instruments; pre- and post- TRACE proficiency scores. Since reliability of the pre- and post- TRACE exams were low and correlations between sections of the exam revealed low and negative correlations readers should be cautious about findings.

Finally, the analyses would have been more sensitive to relations between proficiency test scores and students' perceptions obtained through questionnaires if it was possible to link individual students' proficiency test scores to their response given in the questionnaire.

1. 7. Organization of the Study

The present study comprises five chapters. In Chapter I, a general introduction of the study in which the background information, the statement of the problem, the purpose of the study, the research questions addressed in the study, the overview of methodology (participants, setting, data collection instruments and data analysis), contribution of the study and limitations of the study are outlined.

Chapter II presents review of literature into washback research and consists of headings such as definition of washback, washback models and main findings, factors mediating washback, working for positive washback, and integrated tests as a means for engineering washback.

In Chapter III methodology of the study is presented. This chapter gives information as to setting, participants, mixed-method design, data collection tools, data analysis procedures, piloting study, and validity considerations for the research study.

Chapter IV discusses the results of the research study under three main headings, namely findings related to washback, findings related to instructional sensitivity, and findings related to consequential validity.

Finally, in Chapter V, conclusions and discussions related to the findings, implications, limitations and suggestions for further research are presented. At the end of the dissertation, references and appendices are enclosed.

1. 8. Definitions of Significant Terms

Assessment specifications: are characteristics that prescribe the structure, or organisation, of the assessment as a whole (Bachman and Palmer, 2010, p.371).

Coaching: is any intervention procedure specifically undertaken to improve test scores, whether by improving the skills measured by the test or by improving the skills for taking the test, or both (Messick, 1982, p.70).

Consequential validity: is defined as potential social impact of test interpretation and use (Messick, 1989).

Discourse synthesis: is a constructive meaning-making process of reading for writing (Plakans, 2009, p.563).

English as a Foreign Language (EFL): refers to learning of a language, usually in a classroom setting, in a context where the target language is not widely used in the community (Lightbown and Spada, 2006).

English for Academic Purposes (EAP): is concerned with those communication skills in English which are required for study purposes in formal education system (Jordan, 1997).

Highstakes tests: are the tests whose results are seen-rightly or wrongly- by students, teachers, administrators, parents, or the general public, as being used to make important decisions that immediately and directly affect them (Madaus, 1988, 87).

Instructional sensitivity: is the term which is used to refer to the extent to which student performance on a test or item reflects the instruction received (Kosecoff & Klein, 1974)

Integrative test: is a test in which test-takers are required to use various skills in answering test items as opposed to a discrete-point test in which learners are asked to focus on a single element of language (Davies et al., 1999).

Measurement Driven Instruction: refers to the effect of a high-stakes test of educational achievement on instructional program that prepares students for the test (Popham, 1993).

Washback: is the effect of testing on teaching, and learning (Hughes, 1989, p.1)

Topic familiarity: is the information base that enables test-takers to use language with a reference to the world in which they live in (Bachman and Palmer, 2010, p.41).

CHAPTER 2

REVIEW OF LITERATURE

2.1. Introduction

The review of literature for the present study is comprised of four sections. First, washback is defined and its association to the validity of a test is explored. Then washback models are described and research on how washback operates is outlined. Third, factors mediating different forms of washback and research methods in washback studies are presented. Finally, ways of working for positive washback and using integrated tests as a means for engineering positive washback are explored in detail.

2.2. Washback

Tests impact learning and teaching. Green (2007a, p.1) stated *"washback is grounded in the relationship between preparation for success on a test and preparation for success beyond the test, in the domain to which the test is intended to generalise and to which it may control access"*. These relationships are generally acknowledged to be complex and multifaceted. This section is devoted to different understandings of washback phenomenon manifested in in the literature.

2.2.1. Definition

The terms 'washback' (more commonly used in applied linguistics, language education, and language testing) and 'backwash' (used in general education field) have been used to refer to the effect of testing on teaching, and learning (Hughes, 1989, p.1). Washback has also been formulated as the influence of tests which lead "teachers and learners to do things they would not necessarily otherwise do" (Alderson & Wall, 1993; p.17). For Bachman and Palmer (1996) washback extended beyond the effect of a test solely on teaching and learning. They have associated the concept with test's impact on society, educational policies and individuals since

washback phenomenon is framed as operating at the micro level (i.e. the effect of the test on learners and teachers) and the macro level (i.e. test effect on society and the educational policies and procedures).

Hamp-Lyons (1997, p. 9) also outlined a variety of views put forward by the researchers in defining the concept in empirically based literature on washback and stated that washback has also been conceptualised as a 'curricular alignment' which centres around the interaction between the content of a test and the resulting changes with regard to teaching practices and instructional design. In a similar vein, McEwen (1995 in Cheng & Curtis, 2004, p. 3) drew attention to the notion of curricular alignment and noted that washback has become an increasingly prevalent and prominent phenomenon in education since *"what is assessed becomes what is valued, which becomes what is taught"*.

Many scholars have used different wordings for the concept of washback. A number of labels that are used synonymously with backwash by researchers in applied linguistics entail effects of tests on teaching" (Davies et.al., 1999), on learning (Hughes, 2003), on teachers and learners (Bailey, 1996, 1999; Messick, 1996; Shohamy, 1992, 1993, 2001), on the program (Bailey, 1996, 1999) and on intended curriculum innovations (Cheng, 1997, 2005). While, in the field of educational measurement the washback phenomenon was defined as 'curricular alignment' (Madaus, 1988; Smith, 1991), teaching to test (Madaus, 1988), 'measurement-driven instruction' (Popham 1987), 'test impact' (Baker, 1991; Bachman & Palmer, 1996; Shohamy, 2001), 'test influence' (Alderson & Wall, 1993), 'consequential aspect of validity' (Messick, 1989, 1996), and 'systemic validity' (Fredericksen & Collins, 1989).

Although the influence of exams on teaching and learning has been conceptualised as "backwash" (Hughes, 1989), "washback" (Alderson Wall, 1993) and "impact" (Wall, 1997), Alderson (2004, p115) commented on the backwash versus washback nomenclature, and clarified the distinction between washback and backwash by noting 'there is none'. My own position is aligned with that of Alderson and Wall (1993) who indicated that there is no

difference between these terms on the semantic and pragmatic level. As for the difference between 'washback' and 'impact', scholars have offered different conceptualisations. While Wall (1997) indicated that impact encompasses wider influences of tests beyond teaching and learning, Andrews, Fullilove and Wong (2002) did not differentiate these concepts, arguing that the term washback may include both narrow and wider effects. However, it is commonly acknowledged that the concept of washback is placed within the scope of impact (Bachman and Palmer, 1996; McNamara, 1996; Hamp-Lyons, 1998; Shohamy, 2001, Green 2003). In other words, while the influence of tests on teaching, teachers, teaching methodology, and learning are codified under the term 'washback' (Chen, 2002; Hughes 2003), the wider influences of tests are conceptualised as 'impact' (Bachman and Palmer, 1996; McNamara, 1996; Hamp-Lyons, 1998; Shohamy, 2001, Green 2003). Therefore, impact studies have examined 'macro' issues of test influence including test use, and the social effect focusing on administrators, high-stakes test providers, teachers, parents, and publishers, while washback studies have analysed the influence of tests at a 'macro level' focusing on learners and teachers (Hawkey, 2006, p.7). For Bailey (1996, 1999) washback at micro level comprises of 'washback to the learners' (effects on learners) and 'washback to the program' (effects on teachers, material writers and administrators).

Additionally, Bailey (1999) added that concept of washback entails the effect of a test on further test development endeavours. She argued that with the advent of communicative testing, washback can be utilised as a criterion for both developing and evaluating language tests since a good test encompasses the key characteristics of reliability, validity, practicality and instructional value, which is closely associated with positive washback (Oller 1979 in Bailey 1999), and working for washback should be a tenet for communicative test design, so that communicative tests would explicitly be designed to bring about a positive test influence (Canale & Swain, 1980; Green, 2007a; Hart, Lapkin & Swain 1987 in Bailey 1999).

In this research study, the term 'washback' is used to cover the influence of the institutional proficiency test, the Test of Readiness for Academic English (TRACE), on language learners, learning, teachers and teaching. Accepting a relatively narrow definition of washback, this study locates washback under the umbrella of impact and pursues the exploration of washback of TRACE and its consequential validity.

2.2.2. Washback and test validity

The concept of washback has been debated with respect to its relation to validity. Some scholars (Alderson & Wall, 1993; Davies, 1997) argued that washback should not be regarded as an aspect of validity of a test because there may be other factors which may influence washback. It is argued that a range of social and educational factors (e.g. individual characteristics of teachers and learners and how teachers teach-how learners learn) may affect washback and these factors at work are difficult to spot and beyond scientific measure. Therefore, washback was not incorporated into the concept of test validity. In support of this claim Alderson and Wall (1993, p.116) see no direct relation between validity and washback: *"Whereas validity is a property of a test, in relation to its use, we argue that washback, if exists-which has yet to be established- is likely to be a complex phenomenon which cannot be related directly to test validity"*.

On the other hand, according to Messick (1996, p. 254) washback is an elemental part of construct validity which is conceptualised under consequential validity as an inherent quality of any kind of assessment, especially when the results are used for important decisions. Advocates of Messick's views (e.g. Morrow, 1986; Weir, 1990; Shohamy et al., 1996) also concurred that the effect of a test on learning and teaching is an elemental aspect of its validity. For Messick, washback was associated with the consequential aspect of construct validity: *"In the context of unified validity, evidence of washback is an instance of the consequential aspect of construct validity, which is only one of six important aspects or forms of evidence*

contributing to the validity of language test interpretation and use". Messick classified six distinct aspects of construct validity namely; the content aspect, the substantive aspect, the generalisability aspect, the external aspect, the structural aspect and the consequential aspect. Washback is seen as an instance of test validity and located under consequential aspect because this aspect entails both educational and social consequences. Messick commented that the consequential aspect of validity involved evidence and rationale for evaluating the intended and unintended consequences of interpretation and use of scores *"...in both the short- and long-term, especially those associated with bias in scoring and interpretation, with unfairness in test-use, and with positive and negative washback effects on teaching and learning"* (1996, p.251).

Additionally, Messick marked the necessity of recognising different factors, such as poor and good educational practices, which may influence washback regardless of the quality of the test. Messick (1996) concurred that if a test is not valid due to construct underrepresentation or construct-irrelevancy, then it cannot exert a positive influence on instruction and bring forth good teaching. Conversely, if a test is valid but there is poor teaching, then negative washback cannot be associated with the test. Therefore, only tests which contribute to validity by minimizing construct underrepresentation and construct irrelevancies can bring forth positive washback.

2.3. How washback operates: Washback models

A number of hypotheses and theoretical models have been proposed in order to refine the washback construct in empirical examinations. In the remainder of this section the following models are described:

1) Washback Hypotheses (Alderson & Wall)

2) Hughes' Model

3) Basic Model of Washback (Bailey)

2.3.1 Washback Hypotheses by Alderson and Wall

In their seminal work *"Does Washback Exist?"* Alderson and Wall (1993) specified fifteen hypotheses regarding feedback, and conceptualised the 'Washback Hypothesis'. The posited hypotheses as a result of their own investigations in Sri Lanka and of review carried out by Khaniya (1990), Hughes (1988) and Wesdorp (1982 cited in Bailey, 1999) state that:

1) A test will influence teaching

2) A test will influence learning

3) A test will influence what teachers teach

4) A test will influence how teachers teach

5) A test will influence what learners learn

6) A test will influence how learners learn

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

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

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

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

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

12) Tests that have important consequences will have washback

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

14) Tests will have washback on all learners and teachers

15) Tests will have washback effects for some teachers and some learners, but not for others (1993; p. 120-121).

Although Alderson and Wall's washback hypotheses entail some general (i.e. the first hypothesis is 'a test will influence teaching') and some more specific concepts (i.e. the fourth hypothesis is that a test will influence the rate and sequence of teaching), they did not provide

precise explanation regarding what constituted washback. These hypotheses draw attention to the scope of intended washback and indicate ways how washback can be made explicit. Despite the lack of detailed scope of the washback system within the framework, these hypotheses are of vital importance because they shed light on how to set research into washback, laying out the territory for further studies.

2.3.2. Hughes's washback trichotomy

Hughes (1994) responded to the Washback Hypotheses posited by Alderson and Wall arguing that constituents of washback mechanism should be identified precisely and defined in further detail. He suggested a model which outlined his categorisation of the test effects. Hughes (1989, p. 1) presented that "the effect of testing on teaching and learning is known as backwash" and in his model offered a detailed framework for washback which included description of how washback could be measured and positive outcomes could be pursued. Components relating to the system of washback involve participants, processes and products. Participants refer to stakeholders such as students, teachers, administrators, materials writers and publishers whose perceptions related to the teaching and learning process may be influenced by an exam. Process refers to endeavours in teaching and learning such as materials development, syllabus design, modifications in instruction and methodology, and use of learning and/or test taking strategies. Finally, product encompasses learners' intake, skills and quality of learning (in Bailey 1996, p. 262). Green (2007a, p.78) noted that Hughes' distinction between test effects on participants, processes and outcomes highlights that the influence of a test on participants (the teachers, learners and materials writers preparing for a test and the perceptions and attitudes they bring to the task), triggers modification of their processes (teaching and learning behaviours) and, consequently these impact products (learning outcomes including knowledge of target skills and test scores).

In addition, for Hughes (1989 in Bailey 1996) the means of promoting beneficial washback effect involved:

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

2. Sampling widely and unpredictably.

3. Using direct testing.

4. Making testing criterion-referenced.

5. Basing achievement on objectives.

6. Ensuring that the test is known and understood by students and teachers.

7. Providing assistance to teachers if need be.

Hughes' model was significant because it was the first to outline constituents of the washback mechanisms and conceptualise the phenomenon within a system level. His approach also indicated areas in washback research which warranted analysis.

2.3.3. Bailey's basic model of washback

Building on Hughes' trichotomy (participants, processes, products), Bailey (1996) proposed a basic model of washback. Bailey's Washback Model is built upon similar premises as outlined in Figure 1 below. However, in comparison to Hughes' model, there are multi directional relationships between constituents and participants which extend beyond teachers and students and involve other stake holders such as materials writers, curriculum designers and researchers. In Bailey's model a test exerts effects on the products through the participants and the processes they engaged in while the participants and processes also have an impact on the test since they provide feedback. This mutual interaction between different constituents is displayed through the dotted lines in Figure 1 below. Unlike the Alderson and Wall's Washback Hypothesis, proposing a linear relationship between tests and teaching or learning, Bailey's (1996) model emphasises the importance of the interaction among these different components.

Also, in comparison to previously suggested models, Bailey offered a more detailed description of products.
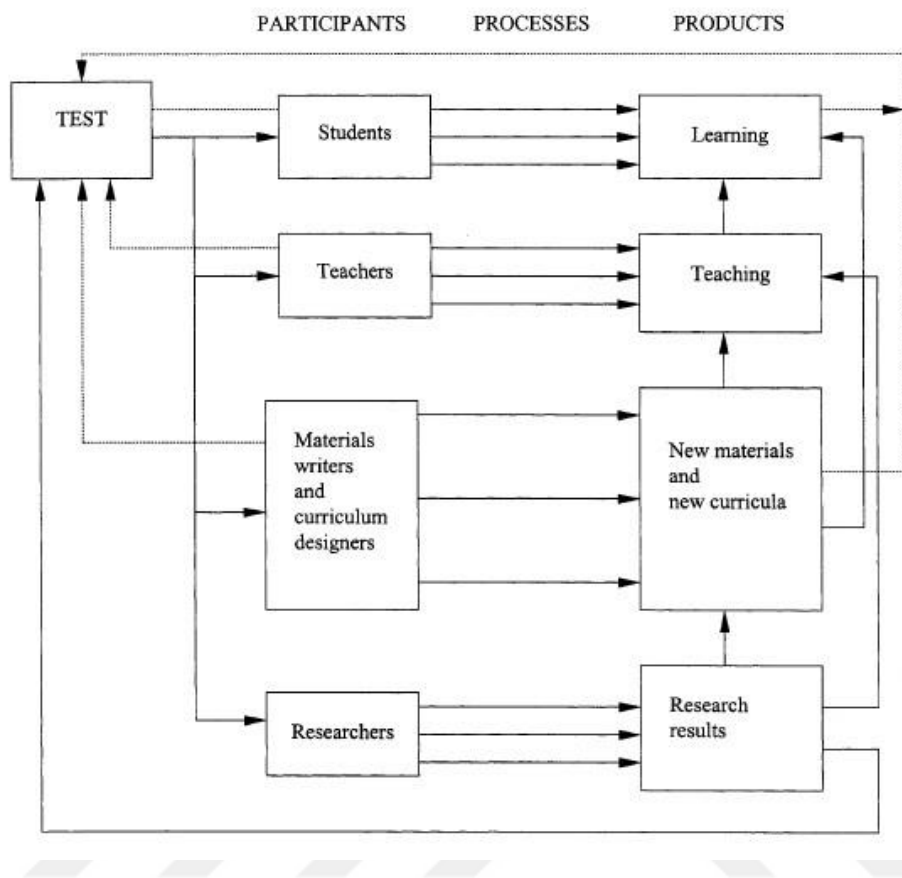
Figure 1: A Basic Model of Washback (reprinted from Bailey, 1996, p. 11)

Bailey further proposed differentiation between 'washback to learners' and 'washback to the programme' and classified five of the hypotheses (2, 5, 6, 8 and 10) put forth by Alderson and Wall under washback to learners. For Bailey some of the processes that students may engage in when preparing for important tests involved; 1) practicing items similar in format to those on the test, 2) studying vocabulary and grammar rules, 3) participating in interactive language practice (e.g., target language conversations), 4) reading widely in the target language, 5) listening to non-interactive language (radio, television, practice tapes, etc.), 6) applying test-taking strategies, 7) enrolling in test-preparation courses, 8) requesting guidance in their studying and feedback on their performance, 9) requesting or demanding unscheduled tutorials or test-preparation classes (in addition to or in lieu of other language classes) and 10) skipping

language classes to study for the test (ibid: 13-14). Washback to the programme, on the other hand, was defined as *"...one result of test-derived information obtained by someone professionally connected with a language program"* (ibid: p.17) and was associated with six of the hypotheses in Alderson and Wall's washback model (1, 3, 4, 7, 9 and 11). However, Bailey did not provide precise explanations as to what kind of processes the participants engage in.

It is important to note here that although both Hughes and Bailey have pointed out that processes affect products, both of the models are devoid of detail as to specification of processes that results in changes in the products. In order to examine the overall instructional effect and account for how much learning had taken place, in this research study the concept of washback on product has been used interchangeably with instructional sensitivity which was believed to be a critical psychometric property of criterion- referenced testing and assessment (Haladyna & Roid, 1981; Polikoff, 2010).

The term instructional sensitivity emerged in the early days of the development of criterion referenced testing when a number of authors pointed out the distinctions between norm- and criterion-referenced assessments (cited in Polikoff, 2010 Cox & Vargas, 1966; Glaser, 1963; Haladyna & Roid, 1976; Popham & Husek, 1969). It was argued that the former was used to show the relative standing of individuals in a group rather than indicating an absolute level of achievement. "Rather, they were designed to maximize variability among test takers in order to differentiate them" (Polikoff, 2010). The latter, on the other hand, was designed in order to measure individual mastery over a particular domain (in Polikoff, 2010 Glaser, 1963). Rather than differentiating among individuals, criterion-referenced tests were intended to differentiate between successive performances of one individual.

Consequently, distinct nature of these two different assessments called for different item selection techniques because item statistics used for traditional norm-referenced assessments such as item total correlation, discrimination index and item difficulty, were not appropriate for

criterion referenced assessments (Haladyna & Roid, 1981; Polikoff, 2010). Psychometricians (in Polikoff, 2010 Cox & Vargas, 1966; Haladyna & Roid, 1981; Kosecoff & Klein, 1974; Popham, 1971; Popham & Husek, 1969) agreed that traditional norm-referenced item selection techniques created certain shortcomings for the criterion-referenced tests. To cope with these shortcomings, they set out to create indices that were appropriate for the criterion-referenced tests with regard to identifying how much learning took place for some individual (in Polikoff, 2010 Brenna, 1972; Cox & Vargas, 1966; Helmstadler, 1972; Kosecoff & Klein, 1974; Popham, 1971; Roudabush, 1974). "This property of items and assessments was called instructional sensitivity" (Polikoff, 2010).

The term instructional sensitivity is used to refer to: the extent to which student performance on a test or item reflects the instruction received (Kosecoff & Klein, 1974), the ability to detect differences in the quality of instruction (Popham 2007), the content and quality of instruction (cited in Polikoff, 2010 D'Agostino et al., 2007; Haladyna & Roid, 1981; Kosecoff & Klein, 1974; Muthen, Kao & Burstein, 1991) and the ability to detect differences in instruction received by students (Polikoff, 2010). Some other terminology was enshrined, with some controversy, and used interchangeably involving "instructional validity", curricular sensitivity", "curricular validity", instructional alignment, and opportunity to learn (OTL)". However, focusing on the test and /or test item rather than the school or teacher, instructional sensitivity is based on an overall instructional effect (Polikoff, 2010).

Instructional sensitivity was examined in the 1960s and 1970s with the prevalent use of criterion referenced testing. It was believed to be a critical psychometric property of assessments (Haladyna & Roid, 1981; Polikoff, 2010). However, although criterion-referenced assessment gained momentum and became widespread, instructional sensitivity became assumed rather than studied (Polikoff, 2010).

Alderson and Wall's Hypotheses, Hughes' trichotomy of washback and Bailey's basic model offer a blueprint for washback research. Bailey's Model for washback set the theoretical framework for investigating the effect of the topic-based language proficiency test, TRACE, on participants (instructors and test-takers), process (teaching involving methods, materials, and tasks) and instructional sensitivity (products).

2.4. How washback operates: Main Findings

2.4.1. Positive and negative washback

Washback has been associated with two dimensions with regard to direction; positive and/or negative (Buck 1988, Alderson and Wall 1993, Brown and Hudson 2002, Hughes, 2003 in Green, 2007a). According to commentators, washback direction refers to the principle that some effects of a test may be beneficial for the teaching methodology (methods and materials) and the development of the learners' progress and achievement, whereas others may be detrimental.

Existing research reveals that some washback appears to be beneficial where as some bring about negative consequences. Consequently, some scholars have suggested using positive washback effects to engineer curricular and instructional changes and innovations. Andrews et al. (2002), for instance, highlighted the possibility of exploiting the power of high-stakes tests to positive ends in support of curriculum innovation, quoting Elton and Laurillard (1979, p. 100) who argue that "the quickest way to change student learning is to change the assessment system". This approach is known as Measurement Directed Instruction (MDI) and it is associated with Popham who defined Measurement Directed instruction approach as effect of high-stakes testing on the instructional program that the student is enrolled in. In this vein Andrews et al. (2002, p. 209) stated that the use of assessment as a means of promoting curriculum change is widespread in education (James, 2000; Chapman & Synder, 2000), and

also in language education (Swain, 1985; Pearson, 1988; Wall and Alderson, 1993; Cheng, 1997, 1998).

Additionally, Wall (2000) argued that positive washback could be utilized to bring about desired changes in effective teaching and learning. Wall mentioned Wiseman's (1961, p. 159-161) Report of the *Consultative Committee on Examinations in Secondary Schools* which outlined the possible effects of tests that could be exerted on teachers as well as students. The good effects on teachers comprised encouragement for 1) a thorough coverage of their subjects, 2) completion of the syllabus within the prescribed time limits, 3) paying as much attention to low achievers as well as high achievers, 4) benchmarking with other educators and schools with respect to standards that were achieved. The possible negative effects included:

> …encouraging teachers to `watch the examiner's foibles and to note his idiosyncrasies' in order to prepare pupils for questions that were likely to appear, limiting the teachers' freedom to teach subjects in their own way, encouraging them to do the work that the pupils should be doing, tempting them to overvalue the type of skills that led to successful examination performance, and convincing them to pay attention to the `purely examinable side' of their professional work and to neglect the side which would not be tested.

On the other hand, prior empirical studies have documented negative test influences on teaching and learning. Green, for instance, indicated that the perception of the negative effects of tests on teaching and learning was widespread and well-established (Vernon, 1956, Wiseman, 1961, Cronbach, 1963, Kellaghan, Madaus and Airasian, 1982, Madaus 1988, Eisemon, 1990, Khaniya, 1990, Corbett and Wilson, 1991, Haladyna, Nolen and Haas, 1991, Kellaghan and Greaney, 1992, Gipps, 1994, London 1997, Bailey, 1999, Jones et al., 1999, Shohamy, 2001 in Green 2007a, p.6). Citing Smith's longitudinal qualitative study carried out in US primary schools (1991), Green (2007a) reported that the negative washback of tests stems from testing programs which "reduce the times available for instruction, restrict the range of curriculum and limit teaching methods, and potentially reduce the freedom of teachers to teach content or to use methods that are believed to be incompatible with the format of standardised tests" (p.7). Similarly Morrison and Tang (2002) expanded the scope of negative and undesired

consequences of washback and added that testing students' ability to repeat book knowledge and facts would lead to negative effects if left unchecked since testing *"becomes part of a self-defeating dependency culture, a hermetically sealed system in which curricula and testing mutually reinforce each other in producing a low-level, facts-driven curriculum, dangerously didactic pedagogy, rote learning, a distortion of student motivation, and a powerful controlling mechanism on teachers and students, and where students are tested to destruction"* (p.290).

Tsagari (2011) found negative washback effects of The First Certificate in English (Cambridge ESOL) on teachers' perceptions and practices. Participants of the study reported they would employ a more communicatively-oriented methodology, focus on individual students' needs and use more authentic materials if they did not have to prepare their students for the test. "Overall, teachers claimed that what was taught and learned in the exam preparatory classes was dominated by the extensive use of exam preparatory materials and teaching to the test practices in order to meet the exam requirements" (p.437). Based on the results of the study, Tsagari outlined several recommendations for language teachers preparing students for a high-stakes exam so as to avoid misunderstanding of the exam requirements that are likely to lead to negative washback on teaching and learning. These recommendations comprise; 1) teachers' familiarity with the exam requirements, 2) acknowledgement of the aims, specifications and administrative procedures of the exam they are preparing for, 3) critical analysis of and content of the test, textbooks and materials and production of relevant materials to cater for the requirements of the exam as well as the needs of the students, 4) providing students with an orientation into exam requirements, 5) providing students with informative feedback, 6) devising coping strategies with the build-up pressure and student anxiety as the exam draws near, 7) balancing teaching and testing in the exam class (i.e. too much focus on testing can highlight students' inadequacies and have a negative impact on self-efficacy), 8) establishing effective channels of communication. Tsagari also argued that there was a need for the inclusion of teacher education programmes which should inform teachers on how to prepare students for

high-stakes exams with regard to choosing exam-preparation materials, familiarising teachers with the rationale and principles behind various instructional techniques in exam materials and train them how to use these materials. (ibid: 439-340). Tsagari's suggestion for teacher education in assessment literacy in this study focused narrowly on providing teacher with know-how into exploitation of exam materials. On the other hand, highlighting the importance of teacher training in assessment literacy, some scholars focused on raising awareness in educational goals in such teacher education endeavours rather than narrowly focusing on the test.

Cheng (2005) and Qui (2005) drew attention to concerns about negative washback effects on teaching and learning encapsulating teaching and learning to the test, practicing for the test excessively at the expense of narrowing of the curriculum, and downplaying educational goals beyond the test (in Xie and Andrews, 2012, p.2). The focus on acknowledgement of educational goals is of vital importance for generating positive washback. Additionally, Bailey (1996) suggested that the negative conceptualisation of washback stems from external-to program standardised tests, which are alleged to lead to a discrepancy between teaching methods and the content of tests. It can be inferred that the discrepancy between teaching methods and the content of tests stemmed from Furthermore, she questioned the compatibility of communicative language testing and 'measurement driven instruction' leading to the view that testing should be the driving force behind curriculum and teaching and learning. Consequently, it is argued that one common underlying feature of good tests is that a test should reflect and encourage good classroom practice. In other words, communicative tests should work for washback by being explicitly designed to bring about positive feedback (p. 261).

The studies reviewed so far have disclosed that washback research has come up with varying results. There are studies that have found positive washback on teaching and learning whereas there are studies that have indicated negative washback. Consequently, it is generally

acknowledged that a test might generate positive or negative washback on how teachers teach and how leaners learn.

2.4.2. Washback on process

In Hughes' washback model (1994) 'process' was described as participants' actions that contribute to learning. In exploring washback with respect to processes, researchers placed effects of tests on curriculum (e.g. Alderson & Wall, 1993; Cheng, 2005; Lam, 1993; Read & Hayes, 2003) teaching materials (Alderson & Hamp-Lyons, 1996; Read & Hayes, 2003; Qi, 2004; Wall & Horak, 2006) teaching methodology (Burrows, 2004; Saif, 2006; Shohamy, 1993; Watanabe, 1997) and teachers' assessment practices (Alderson & Wall, 1993; Wall & Horak, 2006) under process. The literature in the field of washback research reported contradictory findings indicating that in various educational contexts and situations washback is exerted in different ways on process and in some cases there is no washback. The key issues that have been identified in prior research studies are outlined in the remainder of this section.

Narrowing of Curriculum

Some research studies reported that washback generates narrowing of curriculum (e.g. Alderson &Wall, 1993; Li 1990). Read and Hayes (2004) reported washback effects of IELTS test in a comparative study in New Zealand context. Comparison of IELTS preparation courses, one clearly test-focused and the other with a stronger EAP orientation, revealed that the former caused a narrow focus to curriculum leading students and teachers to delve into practice of the test tasks rather than the development of academic language proficiency. In contrast the latter brought about development in students' general language development catering for a wider range of academic study needs which involved less teacher centered instruction and more communicative activities. Azadi and Gholami's study (2013) confirmed this finding. They explored the possible washback effect of English language tests on English language teaching in high schools in Iran by conducting observations of 30 EFL classes, analysing the content of

4 test booklets and surveying perceptions of 30 teachers as well as students. Data analysis indicated a strong relationship between content of the tests and the classroom teaching. Consequently, the research concluded an overall negative washback effect of the English language tests on teaching materials as well as narrowing of the curriculum to testable skills which were reading and grammar due to abandonment of teaching practices which would cater for the communicative competence.

On the other hand, Wesdorp's study (1982) reported conflicting results indicating that multiple-choice language test that were used as a curriculum innovation did not lead to a narrowing of curriculum. However, care should be taken in interpreting the results of this study since data gathering consisted of only one source of data collection which was interview.

Washback on Materials

High-stakes exams affect teachers' use of materials in a variety of ways. In the available literature research findings were inconclusive whether high-stake exams create positive or negative washback on materials. There are studies that have found positive washback. Watanabe (2000) reported that teachers tried to innovate during exam preparation classes by making use of a variety of self-made materials. Similarly, Lam (1994) reported some innovative use of materials generated to complement assessment-driven educational reform involving the use of teacher-produced authentic materials.

However, some research studies discussed negative washback effect on materials due to teachers' inclination of producing exam-oriented materials which were designed for test preparation and gains in scores in high-stakes exams (Read & Hayes, 2003; Wall & Horak, 2006). Lam also criticised teachers as being *"... 'textbook slaves' and 'exam slaves' with large numbers of the former relying heavily on the textbook in exam classes, and of the latter relying even more heavily on past papers"* (Lam 1994 in Spratt 2005, p.)

Spratt (2005) cited contradictory findings of studies which discuss production and use of materials. Some studies (Read & Hayes, 2003; Cheng, 1997 in Spratt, 2005) indicated that when exams were revised new editions of course books and other exam materials were issued. On the other hand, Shohamy et al. (1996 in Spratt 2005, p.10) proved that proliferation of exam oriented materials was not the case for the low-stakes Arabic exam in their study. Consequently, Spratt claims that if exams are regarded sufficiently high stakes they generate the ample publication of exam-related materials. Furthermore, researchers have noted the relation between time and use of exam materials. Alderson and Wall, 1993 in Spratt 2005, p.12) indicated that as the exam gets closer, teachers employ past papers and commercial exam-related publications more.

Although there are contradictory research findings regarding washback effect on materials it is commonly acknowledged that high-stakes exams have led to changes in teaching materials promptly. Cheng (1997) studied the washback effect of the Hong Kong Certificate of Education Examination in English in Hong Kong secondary schools and reported that washback affected teaching materials quickly and efficiently. Test effects on materials are also identified in the introduction of a new exam or a change in an existing exam (Li, 1990; Fullilove, 1992; Lam 1993, Shohamy et al., 1996; Qi, 2004).

Washback as a lever for educational change

Targeting a positive washback effect with respect to student learning, some countries have attempted at realising educational change in English language teaching and learning educational policies by means of introducing new high-stakes exams or integrating modifications to existing exams. The influence of examinations to bring about educational reforms in teaching and learning has been discussed extensively in language education literature (e.g. Wall & Anderson, 1993; Cheng, 1997, 1998)

Wall (2000) highlighted the importance of examining impacts of tests on the 'products of learning' and suggested careful analysis of tests results to concur "whether students learned more or better because they have studied for a particular test" (Wall, 200, p. 502). Andrews, Fullilove and Wong (2002) examined the effect of addition of oral component to Use of English (UE) in Hong Kong to students' spoken English performance. They concluded that introduction of the oral examination exerted some varied influence on students' performance ranging from improved performance to superficial learning outcomes such as producing memorised phrases. One of the conclusions that is drawn from the study has framed the mediating role played by the teacher and the predictability of the effects of a testing innovation. For Andrews et al. the crux of using washback to engineer a pedagogic change depended on teacher perception and on the content of teaching.

2.4.3 Washback on participants

Participants who could be affected by the tests include teachers, learners, educational administrators, materials writers and publishers (Bailey, 1996). Wall (2000) indicated that tests could not only affect participants' attitudes and the activities they are engaged in but also the amount and quality of learning. Numerous research studies revealed while content -what teachers teach- changes as a result of test washback, methodology- how teachers teach- doesn't change (Cheng 2004, 2005; Green, 2007b; Wall 2005). On the other hand, some researchers have claimed tests affected both 'what' & 'how' teachers taught (Alderson & Hamp-Lyons, 1996; Shohamy, 1993; Watanabe, 1996). Another line of research with regard to teacher washback focused on differences between novice and experienced teachers in terms of how teachers react to the test (Shohamy, 1993; Donisa-Schmidt & Ferman, 1996). Although scholars could not reach an agreement as to whether washback affects how teachers teach or what teachers teach or both, it is commonly acknowledged that teachers play an essential role in determining to a greater or lesser extent whether to allow washback to operate, what areas it should operate in and how (Spratt, 2005).

Washback on Teacher

A number of researchers (e.g., Cheng, 2004, 2005; Green, 2007b; Wall & Alderson, 1993; Wall, 2005) uncovered that while "what" teachers teach (content), changes due to test waschback, "how" teachers teach (methodology) does not change. Wall (2005) discovered that teachers focused on the tested skills involving the reading and writing on the New O level English examination in Sri Lanka. Exclusion of four skills instruction was particularly eminent during the examination preparation period. In addition, in contrast to the expectation of a student-centered approach that the examination was expected to induce, teachers were rather observed to employ a teacher-centered approach. In a similar vein Cheng (2005) disclosed that the new Hong King certificate of Education Exam in English encouraged teachers to prioritize speaking and integrated skills aspects of the course. Consequently, reading aloud task was substituted by role play and discussion. However, despite the inclusion of learner-centered communicative activities, Teacher Talking Time (TTT) still remained a big part of classroom teaching.

Chen (2002) construes that the influence of public examinations on teachers' curricular planning and instruction as 'superficial' as the analysis of teacher perceptions collected through a survey and focus group interviews reveal that the washback may influence teachers about what to teach but not how to teach.

In addition, it is documented that individual teacher factors intervene in positive washback which in return brings about changes in methodology of teachers (Wall & Alderson, 1993; Cheng, 2004). Some researchers have claimed that even if teachers would like to change the way they teach, their professional background, poor training and the inability to adequately grasp how exams work mean that they would miss the opportunity to adopt new teaching methodologies (Wall & Alderson, 1993; Cheng, 2004). Chen (2002) elucidated a variety of factors which affects the degree that teachers perceive the impact of public exams on their teaching. These factors are classified into teacher characteristics (involving teaching

experience, education, in-service training education, perceived professionalism in teaching, perceived importance of the exam, gender and perceived awareness of the exam) and context characteristics (including school type, school location, grade the teacher is teaching, students' perceived learning attitudes, perceived attention from external forces and class size). Similarly, some studies have indicated that the washback effect of new tests which were used deliberately as a lever for educational change to change how teachers teach exerted varying effects based on the teacher involved (Alderson & Hamp-Lyons, 1996; Watanabe, 1996; Burrows, 2004).

In addition to above mentioned teacher characteristics that intervene in how washback operates, another factor at play concerns contextual features. As an example to context effect in shaping washback, Chen quotes Shohamy, et al., (1996) who reported teachers who were teaching the upper-level students in their study were inclined to gear their teaching more exclusively towards the tested skills than lower-level teachers because the students in the upper level were closer to take the exam. *"Similarly, Alderson and Wall (1993) also found out that teachers in the upper grades were more inclined to model instruction to meet the objectives of public exams"* (Chen, 2002, p. 5).

On the other hand, according to some researchers (e.g., Alderson & Hamp-Lyons, 1996; Shohamy, 1993; Watanabe, 1996), tests affected not only 'what' but also 'how' teachers teach. Stecher, Chun and Barron (2004) examined the effects of assessment-driven reform endeavours on the teaching of writing in 70 Washington elementary and middle schools by surveying 277 teachers and analysing test scores as well as classroom practices. Researchers articulated that Washington's education reform widely affected both the content and the methodology employed by the participant teachers with regard to their allocation of time to writing, the emphasis placed on specific aspects of writing, teaching methods, and their students' learning activities. *"In most cases, teachers indicated that they incorporated processes and strategies into their existing teaching practice rather than displace familiar lessons and strategies" (p.68).* However, researchers noted that instructional time was reallocated from non-tested

subjects to the tested subjects and concluded that "tests are driving change more than standards" (p.69).

In Lam's research (1994, p.91) findings revealed that more experienced teachers were significantly more "examination-oriented" in comparison to their less experienced colleagues. Similarly, research by Shohamy (1993) and Shohamy, Donisa-Schmidt, and Ferman (1996) revealed differences between experienced and novice teachers. It was argued that while experienced teachers prioritised materials that would be included on the test and taught towards the test, conversely novice teachers focused on teaching oral language through a repertoire of activities. It should be noted that findings concurred with the assertion that the same test may affect teachers differently due to individual differences pertaining to educational background and teaching experience. However, these studies did not examine the reasons behind individual teachers' different reaction toward the same tests.

Alderson and Hamp-Lyons (1996) and Watanabe (1996, 2004b) uncovered significant variations in the way teachers instructed students before the same exam or exam skill. Some teachers resorted to explicitly traditional grammar-translation approaches to help students to learn the linguistic aspect of English, while others incorporated communicatively-oriented approaches to develop students' real-life language ability. Watanabe indicated that the disparity between individual teachers' instruction may stem from different teaching methods different teaching beliefs, amounts of experience, as well as diverse educational backgrounds (2004b).

While a number of research studies have reported contradictory findings regarding washback effect of a test upon what (content) and how (methodology) teachers teach, Pan (2011) attributed this difference to Hawkey's claim that the distinction between course content and methodology is not always clear cut. However, researchers concluded that although teachers may be affected by the tests at varying degrees and whether the test exerts washback on content, methodology, or both, most teachers had the inclination to 'teach to the test' to

increase success rates in test (Cheng, 2004; Ferman, 2004; Gu, 2007; Shohamy et al., 1996; Wall, 2005).

Manjarres (2009) explored the washback of the foreign language test of the state examinations in Colombia through teacher and student interviews, class observations and content analysis of the official exam papers and documented a general overall positive washback. Participants' response portrayed a prevailing general awareness of the importance of improving teaching and learning English as a foreign language. However, the study reported a specific negative washback effect in terms of a narrowing of the curriculum and teaching towards the test. Manjarres stated that orientation of the classroom practices, including class tests, strongly correlated with the direction set by the exam. *"The scope of both the test and the class goals is quite limiting, because most of the time the students are asked to focus on linguistic features, and any activity that goes beyond this (dialogues, writing, and presentations) is seen as a waste of time".*

Therefore, it can be concluded that research studies have come to contradictory conclusions some findings indicating that the test effect led to positive consequences and promoted instruction of communicative activities whereas some others led to negative consequences and brought about instruction of exam-oriented and test-related activities. Additionally, there have been contradictory conclusions of test effect on teachers since some researchers argued that while what teachers teach (content) changes, how teachers teach (methodology) remains entrenched (e.g.Wall and Alderson 1993; Cheng, 1997) whereas, some others (e.g., Alderson and Hamp-Lyons (1996); Blewchamp, 1994; Stecher, Chun & Barron, 2004; Watanabe 1996) claimed that not only content but also methodology of instruction changes.

Washback on Learners

Previous research into washback in learners have reported a variety of findings regarding washback leading to positive attitudes towards tests (Li, 1990; Read and Hayes, 2003), negative attitudes towards tests (Shohamy et al., 1996), changes in the use of learning strategies (Watanabe, 1992) and lack of influence in the use of learning strategies (Gosa, 2012; Pan & Newfields, 2012; Zhan and Andrews, 2013)

Pan and Newfields (2012) explored the effect of mandated EFL proficiency tests on learners by using an experimental and control group in tertiary level institutions with and without English language proficiency requirement. Researchers contented that "standardized tests are not a panacea that will always succeed in changing students' study habits" since tests do not influence students' strategies for learning English or test preparation. Test requirements did not lead to "studying for the test," that is often reported in examination-oriented societies (Chern, 2002; Lai, 2003; Tsai & Tsou, 2009 in Pan & Newfields p. 119). This finding contradicts the conclusions of some washback researchers (Green 2007b, Shohamy, Donitsa-Schmidt, & Ferman 1996; Tsagari 2009; Xie & Andrews, 2012) who claimed that the examination emerged as a strong motivation leading to studying for the test. In addition, since most of the participants employed the old habits of traditional, non-communicative study, a change in students' learning activities was not observed. Pan and Newfields argued that learner washback is mediated in part by teachers and consequently, student preferences for traditional methods of reading texts and memorizing grammatical rules, vocabulary, and phrases stem from their teachers' inclination for teacher-centered instruction.

Using 106 student diary entries and 30 post diary interviews from 3 informants, Zhan and Andrews (2013) investigated the washback effects of a high-stakes examination (College English Test Band 4) on out-of-class English learning of the Chinese non-English major undergraduate students. Data analysis ascertained that students were more likely to change what

they learned rather than how they learnt. Findings of this study resonates with the conclusions of previous washback studies (Bailey, 1996; Ferman, 2004; Shih, 2007; Green 2007a) in that students attach particular importance to skills that are tested. Citing Cheng (1998) and Andrews (1995) Zhan and Andrews (2013) considered this type of washback as "superficial and "quantitative" since the students seemed to adopt drilling and practising test-type exercises rather than fundamentally changing their learning methods.

Gosa (2012) examined the washback effects of the English component of Romanian school-leaving exam in a diary study and found out that 9 out of the 10 diarists did not prepare for the test since they perceived it as 'easy'. Gosa concluded students' expectations of the exam was the sole driving factor in explaining students' perspectives in teaching and learning activities. The students' perception regarding value of the assessment and beliefs about how well they think they will do on the exam not only influenced their attitudes towards teaching, but also their own learning. Similarly, a corollary of Green's (2007b) study was that goals of individual learners and their understanding of test demands affected their learning outcomes regardless of their choice of course and its content (Xie and Andrews, 2012).

On the basis of the literature considered so far, it can be inferred that there have been few studies reporting verifiable gains in student learning. Some research has addressed the effect of coaching for examinations such as the SAT (Becker, 1990; Johnson et al., 1985), and the TOEIC (Robb and Ercanbrack, 1999) on gain scores. Yet it is not clear if gains in these tests are an indication of language competence or a skill to take tests (Amrein and Berliner, 2002; Klein et al., 2000; Munoz and Alvarez 2010).

2.4.4 Instructional Sensitivity

Another facet of the washback phenomenon includes test effect on learning. Research studies have reached mixed results regarding washback on instructional sensitivity. Several researchers explored washback effects on learning by comparing students' gains in scores when exposed to test preparation oriented courses versus English for Academic Purposes (EAP) oriented courses and concluded that there was limited or no washback on learning.

Hayes and Read (2004) studied the impact of the IELTS test on the way international students prepared for academic study in New Zealand. They compared two IELTS preparation courses in university language centers; one clearly test-focused and one the other with stronger EAP orientation by making use of class observations, teacher interviews, teacher and student questionnaires and pre- and post-testing of the students. Researchers observed less teaching of language, more implementation of tasks under test conditions and teacher control in classes in the test-oriented IELTS preparation course. In contrast, EAP oriented IELTS preparation course hosted even coverage of four skills, less teacher oriented activities, more allocation of time to activities that involved speaking and writing, larger range of tasks, wider range of communicative activities. They found out that the IELTS preparation course that is geared towards the test exerted negative washback effects on the students' learning because the teachers and the students tended to focus on practice of the test tasks rather than improving general language competency. In contrast EAP oriented IELTS preparation course appeared to target development of academic language proficiency in a broader sense, cater for a wider range of academic study needs and facilitate students' general language development. Pre- and post-test which assessed students' IELTS performance did not display any significant difference with the exception of the mean scores for the Listening part of the exam.

Similarly, Green (2007) explored the washback of IELTS academic writing test to learning outcomes by comparing learner performance in several types of courses; one with a

test-preparation focus, the other with an EAP orientation with a specific focus on writing and the third one which combines the two. 476 students preparing for academic study in 15 UK institutions took an IELTS writing test to determine their gains in scores and responded to questionnaires used to capture their experiences during their courses. It was concluded that the research "has cast doubt on the power of the dedicated preparation courses to deliver the anticipated yields" (p.93) since learners exposed to test preparation oriented courses did not improve their scores more than pre-sessional EAP courses with EAP focus.

On the other hand, some studies reported positive washback on learning (Hughes, 1988; Saif, 2006). Saif (2006) investigated the relationship between preparation for the (ITA) test and learning outcomes and concluded that there was positive washback. However, as the researcher stated, it is important to note here that although results of the study indicated that high stakes language tests that address the various needs of test takers and the educational system could lead to positive washback in teaching and learning activities, the test by itself cannot create change in the educational system due an intricate web of other interrelated factors.

Hughes (1988) investigated effects of a new university exam in the Turkish context and concluded that the test led to positive washback, increasing students' performance. However, this conclusion had received criticism since the study was devoid of details regarding actual teaching in the classes (e.g. How teachers taught when they were prepping their students for the new test).

Although one of the early washback studies is conducted in Turkish educational settings by Hughes (1988), there are only a few studies conducted on effects of national tests in the Turkish context. Contrary to Hughes' findings these studies argued that there were negative washback effects on learning. Ozmen (2011) examined the washback effect of Inter-University Foreign Language Examination (ÜDS) on candidate academics in Turkey. Interview data of 12 participants indicated negative washback effect of the test leading to inhibition and avoidance

problems. The results indicate that the exam has both a micro level effect defined as the effect on an individual or a small group of individuals, and a macro effect on a relatively populated group of individuals studying or working at university contexts. An obvious corollary was that participants needed to develop more than what the test assessed and therefore, the test represented an obstacle for their learning.

Akpinar and Cakıldere (2013) focus on the impact of two high-stakes Foreign Language Tests (KPDS and UDS) on receptive and productive skills of 103 academicians in Turkey. Their qualitative study indicated that tests exerted positive impact on the tested skill, reading, whereas negative washback was observed in terms of writing, listening and speaking skills since they are untested. It was concluded that since exams prioritised reading skills and grammar knowledge for gains in scores, participants were highly interested in improving their reading skills.

In addition, Sevimli (2007) and Karabulut (2007, in Akpinar and Cakildere 2013) have studied the washback of the foreign language component of the university entrance examination in Turkey on teaching and learning. Both studies have revealed narrowing of curriculum toward tested skills and an uneven teaching and learning of four skills. Sevimli indicated that the productive skills of speaking and writing and the receptive skill of listening are totally neglected and not tested. She attributed this to the washback effect of the test on the teaching and learning of the students and curriculum. Similarly, Karabulut concluded that students and teachers were focusing more on the tested areas including grammar and reading, ignoring listening, speaking, and writing that are not tested.

2.5. Factors mediating different forms of washback

Several factors have been identified in washback research which intervene in how washback is exerted. In her review of washback studies in English language teaching Spratt

(2005, p. 21) classified factors that influence washback into four main categories which involve the teacher, resources, the school and the exam itself.

Table 1

Factors identified by empirical studies as affecting degrees and kinds of washback

Teacher-related factors	Resource, the school, the exam
Teacher beliefs about: • the reliability and fairness of the exam • what constitutes effective teaching methods • how much the exam contravenes their current teaching practices • the stakes and usefulness of the exam • their teaching philosophy • about the relationship between the exam and the textbook • their students' beliefs	*Resources:* • the availability of customised materials and exam support materials such as exam specifications • the types of textbooks available
Teachers' attitudes towards: the exam preparation of materials for exam classes lesson preparation for exam classes	*The school:* its atmosphere how much the administrators put pressure on teachers to achieve results the amount of time and number of students allocated to exam classes cultural factors such as learning traditions
Teachers' education and training: • teachers' own education and educational experience • the amount of general methodological training they have received • training in teaching towards specific exams and in how to use exam-related textbooks • access to and familiarity with exam support materials such as exam specifications • understanding of the exam's rationale or philosophy.	*The exam:* • its proximity • its stakes • the status of the language it tests • its purpose, the formats it employs • the weighting of individual papers • when the exam was introduced • how familiar the exam is to teachers.
Other: • personality	

- willingness to innovate

Similarly Watanabe (2004, p.24) outlined factors, which overlapped with the ones outlined above but adopted a more holistic perspective, that mediated the process of washback as test factors (e.g. test methods, test contents, skills tested, purpose of the test, decisions that will be made on the basis of the test results, etc.), prestige factors (e.g. stakes of the test, status of the test within the entire educational system, etc.), personal factors (e.g. teachers' educational backgrounds, their beliefs about the best methods of teaching/learning, etc.), micro context factors (e.g. the school setting) and macro-context factors (society where the test is used). Next section examines the factors which bring about different forms of washback.

2.6.1. Test factors

Constituents of test design emerge as prominent factors in washback research. Madaus (1988 in Spratt 2005 p.5), for instance, concurred that "it is testing, not the "official" stated curriculum, that is increasingly determining what is taught how it is taught, what is learned and how it is learned". In this vein, debate focuses on the influence of the test design and criticisms towards standardised testing and its dominant use of Multiple Choice Questions (MCQ). Standardised testing is criticised for fostering superficial and exam-driven learning approaches (Entwisle & Enthwisle 1992; Fredericksen & Collins 1989; Fredericksen, 1984 in Xie and Andrews, 2012). Also, multiple choice tests have come under for particular criticism for assessing knowledge-based memorisation and factual recall, excluding high-level thinking skills (Fulcher, 2000; Scouller, 1998 in Xie and Andrews, 2012), atomising knowledge and encouraging poor teaching practices (Wise, 1985; Resnick &Resnick 1992, Prodromou 1995, Hughes 2003 in Green, 2007b). However, others have rejected the assertions, stating that there is scarce evidence that either item or test format can manipulate the efficiency of instruction (Williams, 1996; Mehrens, 1998 in Green, 2012, p. 11).

As a response to focusing on constraints imposed by MCQ on the content of a test, testers argue that tests are neutral measurement tools and negative washback of a test stems from the misuse of the test. It is argued that a test is powerful due to social and/or political functions that it carries out) and misuse of the test causes a negative impact on learning and teaching (Shohamy, 2001 & Jin, 2006 in Xie & Andrews, 2012).

Xie and Andrews (2012) investigated whether test design and use influence test preparation of learners. The findings of their study indicate that perceptions of test design as well as high stakes uses of the test affected preparation. *"Test takers who endorsed instrumental test uses as short-term goals assigned more importance to test taking; the value attached to test-taking motivated them to engage more in preparation via usage of multiple preparation strategies"* (p.6). In addition, it was concluded that there is a negative relation between endorsement of instrumental test uses and the learners' use of socio-affective strategies for seeking help. For Xie and Andrews, although at different levels and embodying a variety of methods, test design and test use do affect the teaching and learning.

2.5.2. Teacher factors

Findings of washback research on teachers have offered insights into how teacher related factors mediate washback and influence teaching. Washback researchers have concluded that there are a variety of factors that account for why washback influences teachers differently.

Burrows (2004), Cheng (1995), Shohamy et al. (1996) and Watanabe (1996, 2004) have found evidence that teaching beliefs and teaching experience of the teachers is a mediating factor for washback. Lam's (1994) washback study conducted in Hong Kong focused on influence of English exam on teaching practices. It was concluded that there were instructional differences between less experienced teachers and more experienced teachers since the former were inclined to employ different activities to teach oral language ranging from creating

authentic materials from the mass media, producing meaningful learning activities, encouraging student participation, to adopting an integrated approach to teaching whereas the latter tended to make use of test oriented activities. Additionally, Shohamy et al. (1996) found evidence that teachers' experience influenced instructional approaches of more experienced and novice teachers in that novice teachers resorted to communicatively oriented activities whereas experienced teachers focused more on the test and test-oriented materials. In addition, Watanabe (1996) argued that educational background of the teachers was one of the mediating factors why washback affected some teachers but not others. It was concluded that teachers with theoretical linguistics major at postgraduate level differed from other teachers with a B.A. degree from teacher college and/or university or those who obtained a B.A. degree from a general university.

Spratt (2005, p.24) indicated that teachers have an important role in determining "to a greater or lesser extent whether to allow washback to operate, what areas it should operate in, and how" and argued that teachers have significant say in determining types and intensity of washback and can become agents for promoting positive washback. Burrows (2004) examined the washback effect of a classroom-based assessment in the Australian Adult Migrant English Programme and she found that the test effect varied from teacher to teacher based on the teacher's beliefs and attitudes. Depending on how teachers responded to classroom-based assessment in different ways, Burrows (2001, 2004) classified teachers into three types including "resisters", "adopters" and "partial adopters". The compatibility of teachers' educational and philosophical beliefs with the theoretical and educational foundations of the new assessment tool rendered the type of teachers that they were.

In addition, Alderson and Wall (1993) concurred that teachers could be directed towards teaching to the test because they felt guilt, shame or embarrassment of poor results from their students' performance in the exams.

Although numerous studies found out that tests led to narrowing of curriculum, cramming for the test, test-oriented teaching and coaching, several studies, on the other hand, have stated that tests did not bring about changes to teaching to a significant degree due to several teacher factors. Furthermore, Pan and Newfields (2012) classified washback studies which scrutinized teachers' perspectives concerning tests into two categories; 1) tests that promote instruction of test-oriented activities (Cheng, 2005; Green, 2007b; Hayes and Read, 2004) and 2) tests that promote instruction of communicatively-oriented activities (e.g. Lam 1994).

Pan (2011) investigated the teacher washback of mainstream English language proficiency tests such as the General English Proficiency Test that functioned as the English certification exit requirement at some technical universities and colleges in Taiwan. 160 teacher questionnaires, class observations and 25 teacher interviews were used with two groups of technical colleges in Taiwan, one group with and one group without an exit requirement. Data analysis indicated that both groups considered the test factors as the lowest priority in their pedagogy and teaching to the test was not prevalent in schools with an exit requirement. However, Pan noted that teachers at schools with exit requirements utilized test-related instruction and a variety of communicatively oriented activities 10% more than their counterparts. The research concluded that the test alone was not enough to create a strong washback since other factors such as educational background of teachers, teaching beliefs, and students' interests affected teaching practices.

Similarly, Li (2008 in Pan, 2011) states that China English Tests (CET) brought about a minor change to teaching since the test did not change what or how teachers instructed. Li (2008) attributed this weak level of washback to a variety of factors such as teachers' beliefs regarding reliability of the test.

Alderson and Hamp Lyons (1996), and by Watanabe (1996) have found large differences in the way teachers teach towards the same exam or exam skill, with some adopting

much more overt 'teaching to the test' , 'textbook slave' approaches, while others adopted more creative and independent approaches. The researchers in both these studies stress that the variable may be not so much the exam or exam skill as the teacher him and/or herself. They go on to discuss various teacher-related factors that may affect why and how a teacher works towards an exam (Spratt, 2005).

2.5.3. Student factors

Learning process could be affected by a wide range of personal variables including students' age and educational backgrounds. Green (2007a, p.80) cites previous research studies (Skehan, 1989; Spolsky, 1989; Ellis, 1994) which outlined the numerous variables such as amount and type of instruction, learner background (age, social class, first language (L1)), and psychological factors including intelligence, personality, motivation, language aptitude and language learning strategies that could facilitate or retard the rate and degree of second language acquisition and could mediate the washback effect.

Shih (2007) investigated the effects of a graduation exam on 29 Taiwanese university students. The study confirmed previous research findings in that acknowledging the importance of the test does not guarantee that most students will study harder. 58% of the participants considered that the test was important, but few expressed that they spent time studying for it. Shih stressed that extrinsic factors (e.g. personal factors such as students' part-time jobs), intrinsic factors (e.g. students' learning attitudes), and test factors (the immediate importance of the test; the way the test-driven policy was implemented) mediated washback. These studies point out that depending on the prestige and status of the tests (high or low stakes) and student related factors tests can have positive consequences such as exerting a facilitating effect to motivate students or negative consequences such as impeding students' interest in learning.

2.5.4 Micro and macro context factors

A number of researchers have discussed some other factors which include the general social context (Wall and Alderson, 1993) and the setting of the school in which a test is in operation (Hayes and Read, 2004; Watanabe, 2004). For Watanabe micro context factors relate to the ethos of the school setting in which the test preparation is carried out and macro-context factors are associated with the society where the test is used (p. 22).

Wall (2005) has investigated the effects of a national high-stakes exam of English in Sri Lanka. She arrived at the conclusion that characteristics of the educational setting as well as the exam-related factors exerted influence on teaching and learning. Lam (1995) also indicated that school culture and educational policies affect how washback is exerted. In their IELTS washback study Hayes and Read (2004) pointed out that school type was one of the factors which brought about negative washback due to commercial pressure on private schools.

2.6. Research methods in washback studies

The effect of a test on the teaching and learning) has been recognised as a very complex phenomenon (Alderson & Wall, 1993; Bailey 1996; Cheng 2000, Watanabe 2004). Although studies investigating washback have commonly employed questionnaire and observation based case studies of participants and processes (Alderson & Wall, 1993; Shohamy et al., 1996; Ferman, 2004; Qi, 2004; Watanabe, 2004; Cheng, 2005), numerous researchers have suggested employing a multi-layered design approach with the inclusion of multiple perspectives to attest the complex nature of the washback empirically. Manjarres (2009) concurred that washback is a complex notion since it can refer to the effect of an examination in the classroom, school, educational system and also the society. Furthermore, washback is mediated by a number of factors, including the teachers' perception of the test, the status of the test and the subject, the macro context where the examination is used, as well as the purpose of learning the language in the context, among others.

Since washback is conceptualised on a variety of dimensions, for Watanabe (2000) *"...the methodology that attempts to disentangle the complexity has inevitably to be multifarious"* (p.20). Therefore, due to complexities of the washback phenomenon, it is generally acknowledged that washback researchers should take account of the whole context of wherein the test is used while amassing data. To this end, Watanabe (2004) has highlighted the importance of *"examining the tests that are used regularly within the curriculum and which are perceived to have educational consequences".* For Watanabe (2004) these requirements call for qualitative research methodology rather than traditional experimental approach (p.22). Additionally, Wall (2000) stated that few empirical studies relied on survey data or on test scores rather than on direct contact with the classroom (e.g. Wesdorp, 1983; Hughes, 1988; Khaniya, 1990; Li, 1990) and advocated for the use of a multi-layered research design to gather data. Therefore, based on the literature reviewed, it is suggested that washback research should take the people that participate in the educational process, the direct contact with the classroom, the outcomes of teaching learning processes into consideration.

2.7. Working for positive washback

As for achieving and maintaining optimal positive washback, a number of factors have been listed, including congruence between tests and educational goals with respect to a small margin of (if any) discrepancy between activities for learning the language and test preparation activities (Messick, 1996; Bailey 1996 in Hamp-Lyons, 1997), test-task authenticity, increased self-assessment and learner autonomy, and detailed (profile) score reporting rather than single scores (Hamp-Lyons 1997). Hughes argued that positive backwash effect can be promoted by; (1) testing the abilities whose development you want to encourage, (2) sampling widely and unpredictably, (3) using direct testing, (4) making testing criterion referenced, (5) basing achievement tests on objectives, (6) ensuring that the test is known and understood by both

students and teachers, and (7) providing teachers with assistance whenever necessary (in Bailey, 1996, p.258).

In addition, Brown (2000) identified a need to focus on test design and test content strategies to foster positive washback of a test. For Chapman and Snyder (2000) test design can induce positive washback, and improve instruction by fostering impact by design, including careful considerations of item format (multiple choice, short answer question, and extended response), content (topic and skills), level of knowledge called for (retention, understanding or use), complexity (the number of content areas and their interrelationship), difficulty (easy or challenging), discrimination (in terms of set standards of performance), referential source (criterion referenced or norm-referenced), purpose (learner performance, curriculum evaluation, teacher evaluation), and type of items (proficiency, achievement or aptitude).

Munoz and Alvarez (2010, p.47) concurred positive washback may be promoted by (1) informing students of assessment procedures and scoring scales, specifying objectives, and structuring assessment tasks, (2) encouraging both teachers and students to establish a clear connection between educational goals and assessment, (3) using self-assessment mechanisms since they aid learners to take control of the assessment since they may be able to diagnose their strong or weak areas, identify current proficiency level, and become more goal-oriented and more self-directed learners.

Saif (2006) investigated the possibility of engineering positive washback by focussing on factors in the background of the test development process and anticipating conditions most likely to lead to positive test effect. The main goal of the study was to examine how a high-stakes performance test based on the practical needs of teaching assistants in an English medium university in Canada would influence teaching activities and learning outcomes of an ITA program. Based on data collected from different stakeholders through interviews, observation and test administration at different intervals in the training program it was concluded that although depth, extent and direction of the effect differed, the ITA test had some influence on

classroom-related areas such as teaching content, methodology and students' learning. Saif concurred the test had directly and extensively influenced teacher's choice of teaching activities and materials. "Class observations and teacher interviews before and after the ITA course suggested that the teacher's methodology and the choice of class activities were, to a large part, adapted to the contents and goals of the test" (ibid: p.28). However, results of the study cannot be generalised to other settings since "they were gathered at an institutional level" (Tsagari, 2007, p.54).

Ultimately, many factors have been offered to instigate positive washback. However, as this research study set out to explore the relationship between TRACE, curriculum objectives of PEP and target skills within and beyond PEP, it is important to focus on the concept of target language use. As indicated by Green (2007a, p.13): "The better a test represents target skills (whether these are based on a specified curriculum or a target domain), through content, complexity, format, scoring procedures and score interpretation, the more beneficial the washback effect is predicted to be (Messick, 1996)".

2.8. Integrated tests to engineer positive washback effect

Theme Based Integrated Language Proficiency Tests

Yu (2013) asserted that there were various interpretations of integrated assessment and argued that perceiving integrated assessment as an equivalent or subordinate to "integrative test" presented a fundamental definitional challenge. Based on the definition given in The Dictionary of Language Testing (Davies et al., 1999) in an integrative test, test-takers are required to use various skills in answering test items as opposed to a discrete-point test in which learners are asked to focus on a single element of language. Yu critiqued this dictionary entry, stating that anything that is not a discrete-point test is considered as integrative. In other words, the concept of integrative tests is blurry and it ranges from a cloze test which may require not only linguistic knowledge but also guessing meaning from a written text to an extended response writing test in

which test-takers are required to combine language abilities, discourse and strategic competence. Consequently, Yu comments that the terms "integrative" and "integrated" are close to one another if two macro level language skills are required to complete an assessment task.   According to Jennings et al. (1999) a number of language proficiency tests, including CAEL, IELTS, OTESL, with the aim of allowing or denying ESL students into university admission procedures utilize a thematic or topic-based approach in an effort to reflect a more real-life use of language in academic domains (p.427). Contrary to a series of independent discrete-point items, a topic-based test uses a more integrated approach since it provides the test-takers with listening, reading and writing tasks that are concerned with one topic. Test-takers are given an opportunity to construct meaning from the input material in the listening and reading tasks of the exam and integrate information from both listening and reading tasks into their written outcome in the essay writing task. It is argued that for academic settings, topic-based integrated language proficiency tests may be justified in making a stronger claim for construct validity than discrete point tests (Jennings et al., 1999; Farhady & Sabeti Dianati, 2000).

Authenticity

Authenticity of a test is often linked to language learning based on the assertion that assessment is a critical component of not only teaching, but also learning. Authenticity of a test refers to forms of assessment which spur on a relationship between learning, achievement, motivation and attitudes, and instructionally relevant classroom activities (Kohonen, 1997 in Bachman, 2000). For Bachman (2000), authenticity is defined as "the appropriateness of the language user's response to language as communication (p.13). It is also defined as the "degree of correspondence of the characteristics of a given language test task to the features of target language use (TLU) task" (Bachman and Palmer, 1996 in Bachman 2000) and congruence of the test situation and the real life situation that the learner is supposed to master according to the curriculum (Doye, 1991 in Bachman 2000). Clapham (2000) suggested that an ideal EAP test would contain authentic university materials and tasks so that preparation courses would

genuinely be preparing students for the skills which they would need when they embark on their further studies in their departments (p.518).

The endeavour of reflecting good classroom practice in assessment procedures can be actualised through the use of topic-based integrated tests which require the test-takers to indulge in a constructive meaning-making process in the integrated tasks of reading, speaking, listening and writing. Integrated tasks can be more authentic and a better means of approximating the demands of real-life communication in an academic environment. It is asserted that instruction on synthesizing reading and writing is a common assignment in the academic course work in English as a Foreign Language writing courses as well as in target language use situation for some of the placement exams (Braine,1989; Carson, 2001 & Horrowitz, 1986 in Plakans, 2009). It is also suggested that the 'discourse synthesis process', which is defined as a constructive meaning-making process of reading for writing (Ackerman, 1991; Green, 1993; Lenski, 1998; Marsella, Hilgers, & McLaren, 1992 in Plakans, 2009, p.563) in first language (L1) composition theory, may provide a conceptual framework for composing processes in academic reading for writing. Discourse synthesis concept could be useful in clarifying an underlying construct for integrated tasks. For Plakans (2009), the integration of modalities such as listening and reading in assessing academic writing can be a means in improving validity, enhancing authenticity, and providing test-takers with the content on which to write (p.561).

In order to investigate the perceived importance of the authenticity of tests, Lewkowicz (1997) elicited the ideas of a group of test-takers on a relatively in-authentic multiple-choice test and a relatively authentic test which encompassed integrated listening, reading and writing tasks. Data analysis of a questionnaire surveying test-takers' perceptions of these two tests revealed that authenticity of the integrated test was deemed as an important test attribute only by a few students. It was concluded that how students perceived the test tasks is informed by their performance of those tasks as well as their familiarity and understanding of the task demands. Therefore, Lewkowicz (2000 in Bachman 2000) suggested that the importance of

authenticity in language tests calls for further research into the nature of authenticity, and how it can be manifested in test tasks, its impact on test-takers as well as the test-taking process, how it is perceived by various stake holders, and the effect it may have on instruction. This research study attempts to crosscheck the claims that have been made about the importance of authenticity in language tests by examining the criteria that is drawn upon for a topic-based integrated institutional language proficiency test of English for academic purposes (EAP), TRACE, and following the outcomes of test decisions over time.

2.9. Chapter Summary

The literature review in this chapter examined how washback has been conceptualised and outlined body of empirical evidence up-to-date regarding how washback operates, factors mediating washback and the research methodology that is generally used in studies investigating washback.

Alderson and Wall's washback hypotheses, Hughes' trichotomy of washback and Bailey's basic model has offered a blueprint for washback research and Hughes' Model for washback set the theoretical framework for investigating the effect of the topic-based language proficiency test, Test of Readiness in Academic English (TRACE), on participants (instructors and test-takers), process (teaching involving methods, materials, and tasks) and instructional sensitivity (products).

TRACE differs from other standardized high-stakes tests in that it specifically aims at identifying test-takers who have the ability to use English for academic purposes in university classrooms. TRACE consists of tasks that replicate target language use (TLU) domain and therefore investigating its washback is of vital importance for engineering positive washback and maintaining consequential validity.

Additionally, this research study involved the use of multiple perspectives through different data collection instruments and linked together test design, participants, process and instructional sensitivity (product) and viability of test decisions variables to examine test effect through teaching and learning processes to proficiency test scores. This comprehensive approach to washback research has not been thoroughly emphasized in the current literature and this study aimed to fill this gap to investigate the complex nature of the phenomenon and trace evidence for consequential validity.

CHAPTER 3

METHODOLOGY

3. 1. Introduction

This chapter describes the methodology of the study. It will outline a description of the research setting, selection of the participants, research methodology, data collection instruments, research questions, and data analysis. The present study comprises three parts. The first part attempts to answer the following research question in relation to washback effect of TRACE:

1) Is there a potential washback effect of the TRACE exam on teaching (methods, materials, and tasks)?

The second part of the study attempts to answer the following research question in relation to whether TRACE is sensitive to instruction:

2) How does the language instruction program, based on EAP skills, result in gains in scores on the writing, listening, and reading parts of the TRACE exam?

The third part of the study attempts to answer the following research question in relation to consequential validity of TRACE:

3) How effective are the decisions made by TRACE in identifying the language competency required for academic study within this university program over the time when students embark on their further academic studies at their departments?

3. 2. Setting

This study was conducted at the Preparatory English Language Program (PEP) at a foundation university in Istanbul, Turkey. The aim of the PEP is to improve students' general English ability and academic skills to meet the language requirements of their major fields of study. The levels of PEP are aligned with the Common European Language Framework, namely elementary (A1), pre-intermediate (A2), intermediate (B1), upper-intermediate (B2) and advanced level (B2+). The process of assigning students to each level is as follows:

Incoming students are directed into different levels on the basis of their scores on a placement test administered at the beginning of the academic year. Based on the scores of the placement test, students with elementary, pre-intermediate and intermediate levels in English are directed to intensive general English courses. Others are required to take the TRACE which is an institutionalised English language proficiency. Those who score at or above the cut off score of 65 out of 100 on TRACE are directed to their university mainstream courses in their departments. Others are placed in upper-intermediate or advanced levels. Students' performance is evaluated by formative and summative tests as well as portfolio assessment at each level

This study focused on students who were placed in upper-intermediate and advanced levels at the onset of 2014-2015 academic year based on their TRACE scores and were exempted from PEP.

3. 3. Participants

There were four groups of participants in this study involving (1) non-native and native English language teachers who are currently employed full-time by the PEP, (2) EFL students who scored below he cut score on TRACE and were placed in were enrolled in the upper intermediate or advanced levels at PEP (3) freshman students in mainstream courses and (4)

university instructors of mainstream courses in the departments. Each group of informants are briefly explained.

3.3.1. English language instructors of the PEP

This group of participants included non-native and native language teachers who were employed full-time at the PEP and assigned to improve students' general English ability and academic skills that they will need to meet the language requirements of their major fields of study. A total of 21 teachers teaching at the upper-intermediate (10) and advanced level (11) participated in the study. All participants responded to questionnaires and 14 of them (9 upper-intermediate and 5 advanced level teachers) took part in one to one interviews. The participant language instructors had varying teaching experience ranging from two years to more than 25 years. All of the participants were holders of a Master's Degree in ELT.

3.3.2. Students enrolled in PEP

The second group of participants included incoming students who were placed at upper-intermediate (B2) and advanced level (C1) in 2014-2015 academic year based on their TRACE scores within the PEP. I fact, TRACE acted as another placement test. That is, if test-takers scored above the cut score on TRACE, they proceeded to their major field departments. If they scored below the cut off score, on TRACE, i.e. got a score below 65, they we Students enrolled in the PEP received English instruction for 4 months to achieve a good level of general English as well as academic English skills that they will need to further their studies at their departments. This group of participants included a total of 147 students (44 upper-intermediate students and 103 advanced levels) aged between 18 and 23 took part in this study. These students came from different cities of Turkey and they had diverse educational backgrounds. However, since they were required to take mainstream multiple choice exams for admission to their high schools

and university before they come to PEP, it can be assumed that they are familiar with gatekeeper high stakes exams, exam preparation and multiple choice exam format.

3.3.3. University instructors of mainstream courses at the departments

The third group of participants comprised 19 instructors who taught courses in different disciplines such as aviation, business administration and culinary arts to freshman students, upon completion of the PEP, in a variety of mainstream departmental courses.

3.3.3.1. Informants of the interviews

A stratified convenience sample of 19 department teachers recruited for the one to one interviews. Department teachers had teaching experience of 2 to 25 years, and they all had PhD degrees in their fields. They were the faculty members who taught freshmen courses to students when students finished PEP and received a passing score on TRACE. Informants came from English medium instruction faculties including Faculty of Social Sciences, Faculty of Business, Faculty of Architecture and Design, Faculty of Engineering and School of Aviation, and School of Applied Sciences and School of Languages. Table 2 presents the composition of the participants.

Table 2

Informants of interviews with university department teachers

Departments	N
Psychology	2
International Relations	2
Under Graduate English	2
Architecture	4
Mathematics	2
Aviation	1
Hotel Management and Tourism	2
Engineering	3
Business Administration	1

3.3.3.2. Informants of the questionnaire for department teachers

The respondents of the questionnaire were a stratified convenience sample of 17 university instructors of mainstream courses teaching at architecture, psychology, engineering, hotel management, gastronomy, business administration departments and teaching undergraduate English courses. 10 of the participants were male and 7 were female and they had work experience ranging from 1-5 years ($n= 4$), 6-10 ($n=3$), 11-15 ($n=4$), 16-20 ($n=4$) and more than 20 years ($n=2$). All participants were holders of PhD in their fields except for 4 with MA or MS degrees. The respondents have been teaching mainstream courses such as photography, video production, movements in cinema, English courses for the faculty, technical drawing, engineering graphics, construction management, techniques for construction planning and control, computer aided industrial design, design visualisation, science and technology studies, introduction to design, design studio, calculus for engineers, differential equations, linear algebra, freshman physics courses, finance, and financial accounting.

3.3.4. Freshman students

The last group of participants included freshman students who started the upper-intermediate and advanced levels in 2014-2015 academic year, received instruction in English language and academic skills and completed the PEP with a satisfactory score on TRACE. 39 students studying at various departments of the university were recruited based on theoretical sampling. These students, who completed instruction in PEP and began to experience academic demands of their departments were believed to provide useful insights to the correspondence between the sufficiency of instruction at PEP and the demands of their departments.

3.4. Research Design

The present research study adopted a mixed method design approach, seeking cross validation quantitative and qualitative data. For Fielding and Fielding (2008), the purpose of mixed-method research designs involves: (1) triangulation with the aim of convergence, corroboration and correspondence of results from different methods, (2) complementarity referring to elaboration, illustration and clarification from one method with the results of the other, (3) development involving the results of one method informing the other, (4) initiation encompassing discovery of paradox leading to recasting the questions or results of one method with the results of the other, and (5) expansion of the breadth of study by using different methods for different components of the study (p. 558). Similarly, Fielding and Fielding (2008) concurred "when findings from independent methods converge, it is not simply a matter of identifying points of agreement" (p. 558). They argue that conditions under which findings are invariant also have to be identified, failures of invariance should be explained, reasons for the conditions determined. It is asserted that the differences between findings from different knowledge sources can be as illuminating as their points of agreement. Duff (2002) also denied a polarity between qualitative and quantitative research, suggesting that researchers should conduct longitudinal research when possible, elicit participants' perspectives on their own behaviour, use participants who know each other and have some familiarity with the researcher, look for recurrent patterns across different lines of data and provide methodological, analytical and perspective/epistemological triangulation whenever possible (in Hawkey, 2006, p.31).

Since one of the purposes of this study was to explore the washback effect, instructional sensitivity, and consequential validity of the TRACE, multiple sources of data were needed as recommended by a number of washback researchers who pointed out significance of using a range of data collection methods to explore washback in context (e.g. Burrows, 2004; Cheng, 2001; Cheng, 2004 Watanabe 2004). For instance, Scott (2007) suggests that interviews can explore

perceptions of different stakeholders, and capture rich, multi-layered accounts which would provide in-depth insights into attitudes and description of reported practices. For Watanabe (2004) rather than one single population, inclusion of different stakeholders is normal in the research into washback. Watanabe claims, "In this way, an attempt is made to examine washback from different perspectives (i.e. data triangulation), as it may be the case that some aspects of washback exist for learners but not for teachers, whereas other aspects exist for teachers but not for learners" (2004, p.29). However, it is also commonly noted that participant perceptions may not always reveal actual teaching and learning practices. Therefore, conducting classroom observations has been suggested as a means of data collection to probe webs of interrelationships between variables and processes (Alderson &Wall 1993; Hughes, 1994). Alderson and Wall (1993) also suggested 'a more ethnographic approach' to examining backwash via direct classroom observation, and other methods of data collection for triangulation reasons. Almost all data collection methods recommended by scholars are integrated into the research design in the present study. That is, this study attempted to integrate both qualitative and quantitative data obtained from different instruments and procedures and crosschecked for more valid interpretation of the findings. Table 3 illustrates the cross-checks performed in this study.

Table 3

Cross-checks of data collection instruments of the study

Data Collection Instruments	Cross-checks
Instructional Sensitivity of TRACE	Interviews
	Questionnaires
	Results of pre and post TRACE
Washback of TRACE	Interviews
	Questionnaires
	Class Observations
Consequential validity	Interviews
	Questionnaires

The next section will discuss the various data collection instruments employed in this study.

3.5. Data Collection Instruments

Several data collection instrument were used in order to collect appropriate data that would help answer the research questions in this study. For the first research question that addressed the washback effect of the TRACE, data was gathered through questionnaires, interviews and classroom observations. Questionnaires were given to students and language instructors of the PEP. Additionally, classroom observations were conducted in advanced and upper intermediate-levels of the PEP program.

For the second research question that focused on instructional sensitivity, data was gathered from different administrations of the TRACE, reporting pre- and post- language proficiency test scores of 147 EFL students who were enrolled in the PEP. A pre-test was administered at the beginning of a 4-month English language instruction period and a post-test was administered at the end of instruction. In addition, the data collection involved both focus group student interviews teacher and one-to-one teacher interviews as well as the questionnaires which surveyed perceptions of different stakeholders as to how much learning happened.

For the third research question, that addressed the consequential validity of TRACE, department teachers and freshman students were surveyed through questionnaires and interviews. The following is a brief description of the instruments used and procedures followed for data collection.

3.5.1. Questionnaires

Several questionnaires were employed to complement the multidimensional data collection process in this study. There were 4 questionnaires used to survey the perceptions of language instructors and students of the PEP as well as freshman students and their department teachers. Each are briefly explained in the next section.

3.5.1.1. Student questionnaire for washback and instructional sensitivity

In order to ensure validity and a cross checking of responses, structured questionnaires with Likert-type items were administered to 123 students who were in upper-intermediate (*n=42*) and advanced level in (*n=81*) of the PEP. These students completed the questionnaire which was designed to survey their perceptions related to washback and instructional sensitivity of the TRACE. This questionnaire, which focuses on unveiling potential washback effects of TRACE and instruction sensitivity, attempted to obtain answers to the first and second research question. The questionnaire was given to the students upon completion of their level in the preparatory program before they sit TRACE. The questionnaire comprised of three parts which involved;

1) Part 1 English Language Background

2) Part 2 Attitudes to teaching, materials, tasks and TRACE and

3) Part 3 Instructional Sensitivity (See Appendix A).

The first part of the student questionnaire that is used to survey students' English Language Background is adapted from IELTS Impact Study (Hawkey 2006: 56). Part 1 and 2 are also informed by The Language Learning Questionnaires (LLQ) originally developed by Bachman, Cushing and Purpura (1993) and finalised by Purpura (1999, cited in Hawkey, 2006). Part 2 and Part 3 of the questionnaire, intended to collect information on cognitive and affective factors through participants' responses to statements which were devised based on curriculum objectives of the PEP and test specifications of TRACE.

The first part of the questionnaire was about the English language background of the participants. The second part attempted to elicit information on 'Attitudes to Teaching Materials, Tasks, and TRACE'. This section consisted of 91 items and yielded high reliability coefficient ($\alpha = .96$). The last part of the questionnaire surveyed participants' perceptions on how much they felt they learned (instructional sensitivity). This section involved 43 items and

indicated high reliability coefficient (α =.86). Overall, the questionnaire consisted of 134 items and yielded a high reliability coefficient (α = .96).

The questionnaire was distributed to participants in a hard copy format and was accompanied by (1) a cover letter which explained the purpose of the research and (2) consent form. Before students responded to the questionnaire, their instructor made them watch a video recorded by the researcher to ensure that they understand the procedures. The video explained the aims of the research and invited honest response from students to increase reliability of the survey results by assuring them. that their response would remain confidential. The questionnaire appeared to have a good internal consistency with The Cronbach's alpha coefficient of .97. Table 4 below presents the reliability coefficients of the sections of the questionnaire.

Table 4

Reliability estimates of the different sections of the student washback questionnaire

Sections of the Questionnaire	Number of Items	Cronbach's Alpha
Part 2 A Questions 1-5	5	.88
Part 2 B Questions 1-19	19	.85
Part 2 C Questions 1-4 (Reading)	16	.86
Part 2 C Questions 5-8 (Listening)	16	.94
Part 2 C Questions 9-12 (Writing)	14	.67
Part 2 C Questions 13-15 (Listening)	13	.89
Part 2 C Questions 16-19 (Grammar & Vocabulary)	8	.91
Part 3 Questions 1-12 (Reading)	12	.90
Part 3 Questions 13-23 (Listening)	11	.99
Part 3 Questions 24-38 (Writing)	15	.76
Part 3 Questions 39-43 (Language)	5	.87

3.5.1.2. Teacher questionnaire for washback and instructional sensitivity

Language instructors of PEP were surveyed in order to harness their perceptions and insights in association with potential washback and the instructional sensitivity of TRACE. 21 language instructors teaching advanced level (*n*=11) and upper-intermediate level (*n*=10) responded to a structured questionnaire with four items on a Likert type scale (see Appendix B). The questionnaire was divided into two sections. The first section aimed at exploring teacher perception of potential washback of TRACE on teaching methodology. The second section surveyed their perception regarding the extent of students' learning. The response to the teacher questionnaire shed light on attempts to explore answers to the second and third research question. Tale 5 presents detailed information about teacher participants.

Table 5

Informant profiles of teacher questionnaire for washback

		Upper-int Teachers *n*= 10	Advanced Teachers *n*= 11
Gender	Male	3	4
	Female	7	7
			1
Major in B.A.	TESOL	4	2
	Literature	5	5
	EFL		2
	Applied Linguistics		1
	Education	1	
	Other		
Major in M.A	TESOL		2
	Literature	1	1
	EFL	5	1
	Applied Linguistics	1	1
	Education	3	1
Years of work experience	1-5	3	1
	6-10	2	6
	11-15	5	1
	16-20		1
	More than 20 years		2
Native Language	Native speaker of English	1	1
	Non-native speaker of English	9	10

Highest	M.A.	9	11
Qualification	PhD	1	
Achieved			

The overarching aim of this questionnaire was to investigate ELT teachers' perceptions in association with; (1) how TRACE affects students' learning and their teaching methodology and (2) how much they think their students learned upon completion of their instruction in the PEP. There were 134 items and each item in the scale was accompanied by a 4-point Likert scale ranging from 'strongly disagree' (1) to 'strongly agree' (4) except for in Part 3 where choices for items 16 to 19 ranged from 'never' (1) through 'always' (4).

Questionnaire was given to participating language instructors in a hard copy. Before distributing the questionnaire, the researcher attended the staff meeting of the participants and explained the purpose of the research. The questionnaire had a short introductory letter and asked for their consent via a consent form. Out of 25 questionnaires that were given, 21 teachers returned their questionnaire. The questionnaire appeared to be satisfactory by showing a reliability estimate of α=0.91.

3.5.1.3. Teacher questionnaire for consequential validity

University instructors of mainstream courses were surveyed through a questionnaire in an attempt to answer the third research question regarding how effective the decisions made on TRACE scores were in identifying the language competency required for academic study within this university program. The aim of the research was explained to the participants and they were asked for their consent through a cover letter that accompanied the questionnaire. The first part of the questionnaire was to elicit their background information regarding their major, work experience, gender, highest qualification achieved, and the courses they delivered at the university. The second section surveyed their perceptions regarding how well PEP prepared the students in English for Academic Purposes (EAP). They were required to reflect their points of view by responding to 12 items in a 4-point Likert scale ranging from *'not prepared'* (1) to *'well-prepared'*

(4). Also, after piloting the questionnaire through read-aloud protocol with one of the university instructors, necessary modifications were done and ambiguous items were reworded. In the third section participants were asked to write their comments about the strengths and weaknesses of the students who completed PEP in writing, reading, speaking and listening skills as well as their grammar and vocabulary in the provided table. They were also asked to provide further comments if they felt necessary. Finally, in the fourth section, participants were asked about their suggestions to improve the PEP.

The questionnaires were e-mailed to 35 university instructors of mainstream courses and 17 of them completed this questionnaire (See Appendix D). The Questionnaire was given in the spring semester in May when prep school graduates had been taking their courses for nearly 5 months. Cronbach's alpha reliability coefficient for internal consistency of the items for the 12 items on the scale in section 2 was $\alpha=0.92$.

3.5.1.4. Questionnaire for freshmen students on consequential validity of TRACE

In an attempt to examine viability of TRACE based test decisions over time, a questionnaire was prepared and given to 39 freshman students who received instruction at PEP and had completed upper-intermediate and/or advanced Levels. These students were given questionnaires towards the end of the spring semester after they took departmental courses and were experiencing academic demands in their further studies.

The questionnaire surveyed (1) information as to their department, (2) their perceptions regarding the extent to which PEP prepared them for their departments in terms of English language, (3) their strengths and weaknesses and (4) suggestions to improve PEP.

Further information regarding the majors of participating students are outlined below.

Table 6

Profile of respondents of freshman consequential validity questionnaire

Department	Number of respondents
Architecture:	4
Interior Design	4
Management Info Systems:	1
Industrial Engineering	6
Computer Science	3
Mechanical Engineering	3
Hotel Management	5
Civil Engineer	1
Psychology	1
International Relations	1
Economics	1
Entrepreneurship	1
Business Management	1
Banking and Financial Management	2
International Business and Trade	2
Business Administration	1
Gastronomy	2

The purpose of the questionnaire was to survey the freshman students' perceptions of the effectiveness of the TRACE-based decisions in identifying the language competency and academic skills required for academic study within their department (See Appendix E). Accompanied with a cover letter and a consent form, the questionnaire gathered information about student perceptions as to how well they think the PEP prepared them regarding their English language ability and academic skills training. This questionnaire had 12 items on a 4 points-Likert scale ranging from *'not prepared'* (1) to 'well-prepared' (4). Participants were instructed to check the appropriate box on the questionnaire that reflected their point of view. Cronbach's alpha reliability coefficient for this part was $\alpha=0.88$.

In addition, freshman students who completed PEP were required to respond to an open-ended question and comment on the strengths and weaknesses of the students who completed PEP. The final item of the questionnaire was intended to elicit their suggestions to improve PEP in order

to prepare students for the use of English for Academic Purposes at mainstream departmental university courses.

The questionnaire was administered in May in the spring semester after they had spent a semester taking departmental courses. Consequently, their observations and experience over the time- in the course of 4 months- were expected to contribute to the research question about viability of test decisions made on TRACE scores.

3.5.2. Interviews

In an attempt to answer the research questions related to the washback effect of the TRACE on teaching (methods, materials, and tasks) and learning, its instruction sensitivity and viability of test decisions over time, samples of PEP instructors, PEP students, and instructors teaching mainstream departmental courses were interviewed. All interviews were recorded and transcribed.

Students of the PEP were interviewed in focus groups. Focus groups can access uncodified knowledge and can stimulate the sociological imagination in both researchers and participants (Johnson, 1996 in Barbour, 2007). Also they can encourage active group interaction (Barbour, 2007, p. 2) which in return can provide a window to insights of the group members. It is seen as a means to compare and contrast experiences by reconstructing their own narrative from accounts of others (ibid p. 26). Additionally, focus groups "provide an opportunity to generate data that are amenable to analysis within the symbolic interactionist approach, which emphasises the active construction of meaning" (p. 37). In a similar line of thought, Marshall and Rossman (2011) concurred that this method leads an individual to socially construct attitudes and beliefs by listening to others' opinions and understandings in forming their own. For Denzin and Lincoln (2005) group interviews have some advantages over individual interviews since "they often produce rich data that are cumulative and elaborative, (b) they can be stimulating for respondents and so aid in recall, and (c) the format is flexible" (p.705). Despite potential downsides focus group

68

interview is acknowledged as a viable option by qualitative and quantitative research paradigms (Denzin & Lincoln, 2005).

Language instructors of the PEP and university instructors teaching mainstream courses were interviewed on a one-on-one basis. According to Leeuuw, (2008) the main advantage of face-to-face interview is "the availability of an interviewer to structure the interview situation and help and motivate respondents" (p. 317). All language instructors who teach intermediate, upper-intermediate and advanced levels in the preparatory program were surveyed to collect info on their perceptions regarding washback and the instruction sensitivity of TRACE. University instructors teaching mainstream courses were interviewed to unveil their perceptions about the viability of test decisions over time.

3.5.2.1. Interviews with the students of PEP

Focus group interviews in groups of three were carried out with students who were placed in upper-intermediate Level ($n=21$) and advanced level ($n=26$) in the PEP. In total 47 students took part in the interviews. Some of the interviews were done in English and some of them were conducted in Turkish based on preference of the students. Focus group interviews took between 20-35 minutes. Interview questions mainly surveyed students' perceptions regarding instructional sensitivity of TRACE and washback of TRACE in relation to teaching materials and classroom activities. To examine students' perceptions of how much learning happened questions triggering their self-evaluation of reading, listening and writing competency were asked. Students' attitudes to teaching materials and tasks and correspondence between teaching-learning and being successful on TRACE were prioritised in the second part of the interviews. Students were interviewed upon completion of their levels before they take the TRACE exam.

3.5.2.2. Interviews with language instructors of PEP

One on one teacher interviews are carried out with teachers teaching upper-intermediate level (*n=7*) and advanced level (*n=7*) in the PEP. In total 14 teachers participated in the interviews. Respondents were reminded about the scope of the research study. Interviews were done in English and took between 25-38 minutes. Interview questions mainly surveyed teachers' perceptions regarding instructional sensitivity of TRACE and washback of TRACE in relation to teaching materials and classroom activities. In the former area, questions were asked in an attempt to examine how much learning took place in terms of reading, listening and writing according to the respondents. Teachers' attitudes to teaching materials (including course book, supplementary materials, vocabulary and grammar booklet, Blended Learning Web Materials) and tasks and correspondence between teaching-learning and being successful on TRACE were prioritised in the latter. These interviews aim at surveying their opinions regarding the correspondence between objectives of the PEP and the effect of TRACE on teaching and learning.

3.5.2.3. Interviews on consequential validity: Viability of TRACE-based test decisions with the university instructors of mainstream courses

One-on-one interviews were conducted with the university instructors of mainstream courses to harness their perceptions regarding the viability of the TRACE-based decisions over time. University instructors of mainstream courses were asked to make comments about the language competency and academic skills of their students in meeting the expectations of the academic demands of the departments. In total 19 teachers teaching at the departments participated in the interviews. Informants of the interviews were selected based on theoretical and stratified sampling. They were the faculty staff who taught first year students when students finished PEP and passed TRACE. Informants came from faculties which had the medium of instruction in English including Faculty of Social Sciences, Faculty of Business, Faculty of Architecture and Design, Faculty of Engineering and School of Aviation, School of Applied Sciences and School of Languages. Some interviews were conducted in English and some

interviews were done in Turkish based on the preference of the participants. Interviews took 10-35 minutes.

Table 7

Informants of teacher interviews on consequential validity

Departments	N
Psychology	2
International Relations	2
Under Graduate English	2
Architecture	4
Mathematics	2
Aviation	1
Hotel Management and Tourism	2
Engineering	3
Business Administration	1

3.5.2.4 Classroom Observations

Classroom observations have become one of the most frequently used instrument for gathering data in empirical washback studies (e.g. Alderson and Hamp-Lyons, 1996; Cheng, 1996; Huang, 2009, Tsagari, 2007, Qi, 2005; Watanabe, 1996). Direct observation endows researchers with a more accurate viewpoint of what happens during instruction in the classroom by providing them with the opportunity of collecting live data from live situations (Cohen et.al., 2000 in Huang, 2009, p. 97). Watanabe noted that the type of observation instrument varies based on the contextual factors, examination under inquiry and the purpose of the research and in many cases researchers had to devise an observation tool (Stecher, Chun, & Barron, 2004; Sawille & Hawkey, 2004; Cheng, 2004; Qi, 2004 and Ferman, 2004 in Watanabe, 2004, p.30) or modified an already existing instrument. Consequently, in order to effectively capture and record observed data, researchers have been devising a number of classroom observation instruments conducive for repeatability in other teaching contexts (Flanders, 1970; Moskowitz, 1976; Fanselow, 1977; Allen, Frohlich, and Spada, (1984) (in Huang, 2009). Communicative Orientation of Language Teaching Observation Scheme COLT (Spada & Frohlich, 1995) is one of the most widely used and modeled observation tool among washback researchers (i.e. Huang,

2010; Hayes & Read, 2004; Burrows, 2004). COLT has been used to examine the extent to which different language classrooms display the features of the communicative approach to language teaching since it was constructed with the aim of differentiating communicative language teaching from the more teacher-centered and form focused teaching (Huang, 2009, p. 98). Class Observation tool in this study was modelled after COLT Part A in which *"the observer makes a detailed note in real time on the activities and episodes that occur during the lesson, including the time taken for each one"* (Hayes & Read, 2004, p.102).

In order to examine the washback of TRACE, class observations were conducted with upper-intermediate and advanced level classes. There were 13-17 students in classes. Teachers who agreed to participate in class observations were among the 15 teachers who were interviewed during the earlier phase of the research study. Classroom observations were conducted in order to validate teachers' self-reported data through interviews and questionnaires. In other words, observations predicated on the underlying hypothesis that what the respondents had stated about their teaching practices and students' learning would be observable in their actions. Observation instrument consisted of six categories; time allotment, teaching materials, skill focus, activity, student work mode and comments. Under 'activity' I noted the activity used in class as well as how much time was allocated to each activity. 'Activity' was open-ended excluding any pre-determined descriptors for this category. Descriptor of the 'teaching materials' category pertains commercial course books, supplementary materials, self-edited materials, materials from the internet and others. 'Skill focus' category was accompanied by the following descriptors; listening, speaking, reading, writing, integrated, vocabulary and grammar. Finally, 'students' work mode' was described in terms of individual, pair, group and choral.

Before the observations brief pre-observation meetings were conducted with the teachers to obtain information regarding their educational background and work experience. Teachers were also briefed about the aims of the research. During the observations I took notes on the observation note-taking sheet and coded information under the relevant headings (See

Appendix H). All the observations were video-taped. Rapley (2007) says that video data offers certain benefits including a record of non-verbal conduct, and an aid to the transcription process by the display of which a person is saying a specific thing (p. 39). Also recorded videos provided guidance when there was ambiguity regarding the time allocated to different activities recorded on the site. The duration of each activity was later calculated as a percentage of the total class time using the minute as the unit for measurement. One 50-minutes lesson was observed for 15 language instructors teaching at upper- intermediate (*n=9*) and advanced levels (*n= 6*) in PEP; a total of 800 minutes-13, 3 hours.

Table 8

Background information about language teachers who participated in observations

Gender		Major		Years of experience				Highest Qualification	Native	
M	F	B.A	M.A	1-5	6-10	11-15	16-...	M.A 14	NNS NE	N
6	9	Literature 4 EFL 8 Applied Linguistics 2 Education 1	EFL 4 Education 2 Literature 1	2	4	8	1	PhD: 1	13	2

Upon completion of each observation post-observation meetings were held to discuss the rationale behind the variety of teaching activities employed in the class.

The analysis of the classroom observations was conducted both qualitatively and quantitatively and followed a two-step process. First of all, I collated the data retrieved from the coding scheme in the observation instrument which was used in the real-time observation. The data gathered canvassed the general description of the lessons and served as a background for the further analysis of the videoed lessons. In the second step I watched the videos of the observed lessons and confirmed the recorded the length of time allocated to each activity and identified time allotment to teacher and student talk. The unit of analysis was minutes during this process. In addition, I focused on two dimensions of teaching; what teacher taught and how

they taught by analysing elements of classroom activity. Unit of analysis was the classroom activity since classroom activities employed in instruction represent teacher's methodology and content. Elements of classroom communication shed light on methodology employed in the classroom (how) and classroom activities accounted for the content of instruction (what).

When examining classroom activity quantitatively the focus was on frequency of 1) the language skills or knowledge focus of the activity, 2) the students' work mode during the activity and the amount of time allocated to teacher talk and student talk. English medium teaching and learning is expected as institutional policy The occasional student talk in Turkish was not added to the students' talk time.

On the qualitative end of the spectrum, field notes were analysed to recognize differences among what and how teachers taught in daily instruction.

3.5.3 TRACE

PEP students' scores from different administrations of TRACE were used to evaluate instructional sensitivity of the exam. TRACE in September, 2013-2014 academic year, determined language level and enabled placement within PEP. This administration of TRACE was taken as pre-test in the study. Another TRACE, a post-test, was given to the students in January 2014 upon completion of a 4-month English language instruction period. The comparison of students' pre and post-test grades were expected to provide information on whether TRACE is sensitive to instruction by focusing on (1) the language proficiency levels of the test-takers sitting for the TRACE exam and (2) whether the language instruction program, based on EAP skills, results in increased scores in the writing, listening, and reading parts of the TRACE exam.

3.5.3.1 TRACE; Constructs and Sections

The TRACE is an institutionalised English language proficiency test, which has adopted a topic-based approach in an effort to reflect more closely the actual language use in academic

domains. TRACE was designed under the supervision of an internationally renowned testing and assessment expert when PEP was founded eight years ago. Therefore, curriculum and testing coexisted and improved while curriculum and test designers engineered instructional design of PEP. Consequently, curriculum objectives and TRACE are closely aligned. Curriculum objectives focus on four skills as well as grammar and vocabulary. Theoretically, TRACE was meant to be a proficiency test which is independent of any curriculum, instructional materials and teaching methods (Bachman and Palmer, 2010; Davies, 1990). However, since TRACE has been repeatedly administered through years without any modification, one can assume that instructional materials could have been geared towards the content of the test. This might have been a reason for the washback effect observed in this study. Having taken into account the real life academic needs of university students and the belief that the student needs to use the language to communicate actively and effectively, test developers integrated three language skills (reading, listening and writing) both receptive and productive, in TRACE exam so as to reflect an authentic context. Therefore, TRACE is an integrated skills exam that is intended to assess the student's ability to use language rather than memorize formulas and recognize structures. It reflects the principles of the curriculum of the PEP since curriculum objectives and tested skills in TRACE overlap. Language components such as grammar and vocabulary are tested indirectly through reading, writing, and listening The exam is designed taking CEFR B1 to C1 levels as the reference. There are four sections in the exam which include an unassessed introduction part, reading tasks, listening tasks and a writing task. The table below outlines the components of TRACE, weighting of each section in the overall grade and the time allocation.

Table 9

Components of TRACE

Parts of the test	Weighing	Time
Part 1- Introduction	Not assessed	10
Test Format: Note-taking and brainstorming		minutes
Test Content: Visuals, photographs, short sentences, cartoons		(mins)

Part 2- Reading	30%	80 mins
Test Format: Multiple choice questions (MCQ)		
Test Content:		
Section 1: 2 short texts,		
Section 2: 1 medium length text		
Section 3: 1 longer text		
Section 4: comparing or contrasting 2 or more of the texts from sections 1-3 through cross textual analysis reading comprehension questions		
Part 3- Listening	30%	50 mins
Test Format: MCQ, note-taking		
Test Content:		
Section 1: lecture		
Section 2: 1 or 2 short conversations		
Part 4- Writing		
Test Format: Discursive essay writing using sources (ideas from readings in the exam and note-taking sheet from listening)	40%	80 mins

All sections are concerned with one general topic, usually selected from the field of psychology, sociology, environment or business that test takers are assumed to be familiar with. Also, together with the pictures and visuals that are associated with the exam topic in the introduction part of the exam, reading passages and listening texts provide the test-takers with a substantial context. Each sub-section of the TRACE is briefly explained below.

Section 1: Introduction to the test

This section aims at contextualising the test for the test takers and preparing them for the test by (1) providing background knowledge on the topic to the students, (2) allowing students to personalize the topic, (3) encouraging students to brainstorm on the topic, and (4) activating the students' schemata on the subject. Test-takers are required to watch a slideshow that is composed of photos and visuals, and brainstorm about the topic and note down their ideas on outline sheets involving hard-copy of visual materials that they are presented with a given time frame of 10 minutes. At times they are also presented with slogans and/or short sentences accompanying the visuals, which do not demand high level of cognitive load with respect to reading skills. Students'

introductory outlines and their preliminary notes of reflections are given back to the students in the writing section to provide them with a substantial context. This section is not assessed. Introduction to TRACE (Section 1) is not assessed and is hoped to reduce exam anxiety.

Section 2: Reading

This section of the exam has several reading texts and accompanying tasks which aim at testing the ability of students in comprehending main ideas, finding specific information, understanding details, guessing meaning from context, inferring main argument(s), inferring the author's purpose, tone, the intended audience, and source. These skills, which are listed in the test specifications, are assessed in each form of the test. A variety of task types involving multiple choice, matching and open-ended items are utilised. Test-takers are required to read four reading passages and answer questions based on what they have read in 80 minutes. There are four texts of different genres with varying word counts ranging from approximately 250, 750 to 1000. Test-takers are required to respond to 30 questions which examine reading ability. In reading and listening sections test-takers mark their response to multiple choice questions by filling a circle, on their optic forms. For several open-ended questions test-takers are required to write down their response and teachers fill in the optic form based on correct and wrong answers. Then, optic forms of the test-takers are graded through an optical reader.

Section 3: Listening

The listening section requires the test-takers to take notes on the content of a lecture and use their notes to answer the questions, intended to test their listening ability to comprehend main ideas, find specific information, understand details, and infer the tone and attitude of the speaker. In the listening section of the exam, first test-takers take notes and respond to comprehension questions after the lecture listening, and then listen to a conversation and respond to questions while-listening. The lecture has a recognizable organizational structure with explicitly used discourse markers, repetitions, signposts and explanations to help test takers to follow the

organisation and comprehend the content. The lecture is around 900 and 1000 words and lasts for about ten minutes. Test-takers take notes on given outline sheets while listening to the lecture, and then respond to 20 questions that involve multiple choice items, open ended questions, and matching. In the second listening task -while-listening- test-takers are asked to listen to a conversation that lasts for around 5 minutes. They are required to respond to the questions while they are listening. Students have to respond to ten multiple choice questions while they are listening to the conversation. The listening section lasts for 45 minutes in total. This section is also graded objectively as mentioned in the previous section through the optical reader.

Section 4: Writing

The writing section of the TRACE exam requires the students to write an essay of about 450 words. Students are encouraged to synthesize content from listening-to write and reading-to write tasks of the exam as well as their topical knowledge elicited at the beginning of the exam during the non-assessed introduction section. In this vein, at the onset of the writing section, test-takers are given their brainstorming notes taken during the Introduction to the Test, as well as their notes taken during lecture note-taking task in the listening section of the exam.

Writing question is typically of general interest and would not call for any specialised information.

The writing section of the exam tests the ability of the test-takers to respond to the given essay question in a logical and meaningful way. Test takers are required to compose a logical, well structured, and organized essay in an academic style. They need to develop the controlling ideas in a logical and coherent way through a thesis statement, topic sentences, supporting sentences with examples and details. Also the writing should entail a variety of structures and vocabulary relevant to the task. Additionally, test-takers are expected to use the mechanics of writing accurately. Students are given 80 minutes (10 minutes to be suggested to be used specifically for brainstorming on the topic, 60 minutes to be used for writing the essay and a final 10 minutes used

specifically for proofreading). Students are expected to respond in an academic style in response to the given question using different organisational patterns for developing ideas such as compare-contrast, listing, classification, cause-effect, and opinion. Reading and listening sections of the test are scored by language instructors using an answer key for objective marking. In the grading of the writing section, instructors attend standardisation sessions and examine the analytic criterion that is used for scoring the papers. The criteria include grammar, lexis, organisation and content sections.

3.6 Data Analysis

The following section describes the data analysis procedures used in the present study. For the first part, student and teacher questionnaires were administered and interviews were conducted with the both parties to reveal their perceptions regarding the washback and instructional sensitivity of the proficiency exam, TRACE., In order to answer the first and second research questions, the data gathered from the student and teacher questionnaires were analysed using descriptive statistics to examine the nature of the test washback with regard to materials, teaching methodology and instructional sensitivity.

The data gathered via semi-structured interviews carried out with the participants, were analysed using Bogdan and Biklen's (1998) framework. The interviews were transcribed, and then as a result of intensive and repeated reading, conceptual themes pointed out by the recurring words and ideas were sorted. The emerging conceptual categories which led to major themes, were classified under specific headings with the goal of providing examples to answer research questions. Additionally, the results were quantified where possible to get a preliminary overview of data. Finally, detailed explanations were presented in association to the research questions.

In addition, the data was enriched by classroom observations which were conducted through Communicative Orientation of Language Teaching Observation Scheme COLT (Spada & Frohlich, 1995). All classroom observations were recorded. The researcher took field notes. Quantitative and qualitative methods were used to analyse the observation data.

For the second part of the study investigating instructional sensitivity of TRACE, a repeated measures ANOVA was conducted to compare the pre- and post- proficiency scores of the students to identify whether the test was sensitive to instruction. Analysis determined whether there was any improvement in their proficiency related to the three language skills (reading, listening and writing) upon completion of the PEP. Scores were also analysed statistically to find out correlations between different skills.

Finally, to answer the final research question, namely consequential validity, data gathered from the semi-structured teacher interviews were transcribed and coded according to Bogdan and Biklen's (1998) framework and teacher and student questionnaires were analysed statistically.

3.7. Piloting

In terms of development of the data collection instruments and ensuring validity and reliability a piloting study was conducted prior to the data collection.

For accomplishing questionnaire development, several piloting procedures, including readaloud protocols and trialling the questionnaire with a representative sample, were employed. The student questionnaire for washback and instructional sensitivity had been piloted with 66 students enrolled in advanced and upper intermediate levels of the English Preparatory Program in order to gather information regarding the relevance and clarity of the questions, check user-friendliness of the format and determine the amount of time required to respond to the questions. The reliability of the questionnaire and found to be 0.91. Then each part was further analysed

and items were fine-tuned. To illustrate, analysis revealed that Part 2 Section A which consisted of 7 items that investigated perceptions related to teaching, materials, tasks and TRACE had the reliability score of the Cronbach's alpha coefficient .58. Consequently, items which affected reliability adversely were modified and/or removed. At the onset the questionnaire had 141 items and as result of reliability analysis these items were reduced to 138 and some of them were fine-tuned. In addition, the teacher questionnaire of washback and instructional sensitivity was piloted in read aloud protocols with four language instructors. Think-aloud protocols involved participants thinking aloud as they went over and responded to the questionnaire. Teachers were asked to say what they think about the items of the questionnaire and comment on the content and wording of the items. Think-aloud sessions were audio-taped. These recordings were around 35-40 minutes and they enabled the researcher to go back and refer to what participants did and how they reacted. Based on the analysis of participants' response in these video recorded sessions the wording of some items in the questionnaire leading to ambiguity were changed and explained in finer detail. In addition, 'freshman questionnaire for consequential validity' was piloted before administration with 22 freshman students.

For the development of the observation note-taking and analysis tool qualitative input and a pilot observation was used. Qualitative input consisted of theoretical resources from related research studies which also gathered data through observations and interview with a colleague who also watched the videotaped lesson with the observation analysis form. Class observation allowed the researcher to go into one of the classrooms in Preparatory English Program to observe and iron out design issues regarding the content and format of an observation tool. The lesson was videotaped and a colleague who is a PhD candidate in teacher education tried out the observation analysis form which had on-site coding and lesson video analysis. The observation instrument was shaped upon the feedback and suggestions that were revealed as a result of piloting procedures.

In addition, interview questions were also discussed with and PhD candidate whose field of study was also washback. Questions were piloted with teachers and students.

The pilot study was conducted between June and December 2014. Piloting study involved a variety of procedures ranging from qualitative to quantitative to ensure validity and reliability. Qualitative input, and readaloud protocols were utilized so as to certify content validity and reliability whereas trialling of the questionnaires ensured construct validity. Pilot study had been a beneficial and necessary process which allowed important changes to research questions, the methods used and logistical considerations which contributed to the overall quality of my research design and findings on a massive scale. The table below presents the outline of the pilot study.

Table 10

An outline of the piloting study

Piloting Method	Time	What was done	Participants
Trialling questionnaire	June 2014	Freshman students' questionnaire for consequential validity was given to a group of students.	22 freshman students
Readaloud protocols	June 2014	Inviting teachers to read the items on the questionnaire and talk about what they thought about items.	2 university teachers of mainstream courses
Readaloud protocols	August 2014	Inviting teachers to read the items on the questionnaire and talk about what they thought about items.	4 language instructors from Preparatory English Program
Trialling questionnaire	August 2014	Student questionnaire for washback and instructional validity was given to a group of students	66 advanced and upper-intermediate level students enrolled in Preparatory English Program
Qualitative input	September 2014	Analysing theoretical resources from related research studies which also gathered data through observations	1 language instructor from Preparatory English Program
	December 2014	Conducting a class observation Inviting a colleague to watch the videotaped lesson and ask for comments of the observation analysis form.	
Qualitative input	September 2014	Discussing interview questions with teachers	2 Language instructors, 2 university teachers of mainstream courses

| Piloting questions with teachers and students. | 6 advanced and upper-intermediate level students enrolled in Preparatory English Program |

3.8. Validity, reliability and ethical considerations

3.8.1 Considerations for validity

3.8.1.1 Thrustworthiness of data

Guba and Lincoln (1985) discussed the concept of trustworthiness and proposed four criteria for judging the soundness of qualitative research. Reliability and validity of data was maintained according to four criteria they suggested; credibility, transferability, dependability and confirmability. Their 'Trustworthiness Criteria' roughly parallels validity typology of the positivistic research paradigm. "Thus, credibility in qualitative research is said to correspond to internal validity in quantitative approaches, and transferability to external validity or generalizability, dependability to reliability, and confirmability to objectivity" (Morrow, 2005). Guba and Lincoln typology was used as the framework to demonstrate the extent to which the study is trustworthy in terms of qualitative research paradigm.

Credibility

Credibility is one of the components of the trust worthiness typology and it refers to "the match between the constructed realities of respondents (or stakeholders) and those realities as represented by the evaluator and attributed to various stakeholders" (Guba & Lincoln, 1989; 237 in Lynch 1996). Credibility can be maintained through several techniques which involve; prolonged engagement, persistent observation, triangulation, progressive subjectivity, peer debriefing, negative case analysis, member-checking and referential adequacy.

In order to ensure credibility several techniques have been taken into consideration. Using prolonged engagement, the researcher pursued to trace participating students during one academic year not only in PEP but also in their freshman courses. Another way of ensuring

83

credibility is progressive subjectivity which involves keeping field notes and reflective commentaries (Lincoln and Guba, 1989). Therefore, progressive subjectivity enables the researcher frame the findings of the study during the data analysis and interpretation.

Another criterion is member checking which refers to the informal or formal checking of data, analytic categories, interpretations and conclusions with the informants who provided the data (Lynch 1996). In order to fulfil this criterion, language instructors and university teachers were e-mailed the interview transcriptions and were asked for their options to check whether any meaning was lost during the transcription and/or translation process. Also, after classroom observations I conducted brief post observation meetings with the language instructors shared the observation notes as a measure of confirmation.

In addition, triangulation contributed to strengthening the validity of research findings. Triangulation encapsulates the gathering and reconciling of data from different sources through different data collection techniques (Lynch, 1996). Data was gathered from a range of participants in this research involving language instructors of PEP, university instructors, PEP students and freshman students through different data collection instruments in the form of questionnaires, interviews, classroom observations, test scores from different administrations and student sample work.

Peer debriefing was used as another technique to certify credibility and overall validity of the research. Peer debriefing is defined as an extensive discussion between the researcher and a disinterested peer concerning findings, conclusions and tentative hypotheses (Lynch 1996). During analysis of transcripts of interview data, I consulted two academicians who held PhD degrees in ELT to benchmark the conceptualisation of data with regard to meaning coding, condensing the meaning and interpreting. These two other evaluators and I went over three transcripts and coded them individually. Then we compared our coding scheme to identify similarities and differences. To determine reliability number of agreements was divided by total number of agreements and disagreements. The disagreements were resolved in further

meetings. Also, I presented a certain part of the research during the last phases of data collection process at an international testing conference and received insightful comments from experts and other PhD candidates.

Transferability

Transferability refers to the degree to which the results of qualitative research can be generalized or transferred to other contexts or settings. One technique to enhance transferability is describing the research context and the assumptions that were central to the research extensively and thoroughly. In accordance with this criterion a thick description, an extensive and detailed description of the research context, is provided.

Dependability

Dependability refers to "the way in which a study is conducted should be consistent across time, researchers, and analysis techniques" (Gasson, 2004, p. 94). Thus, the process through which findings are derived should be explicit and repeatable as much as possible. To accomplish dependability and enable readers of the research report to develop a thorough understanding of the methods and their effectiveness an audit trail can be provided. Morrow suggested that an audit trail can be in forms of a detailed chronology of research activities and processes, influences data collection and analysis, emerging themes, categories, or models and analytic memos (p.252). To establish the principle of dependability criterion, the study provided information regarding the research design and its implementation, the operational detail of data collection, and reflective evaluation of the process of inquiry undertaken. Also, video and tape recordings, field notes, codes, analysis sheets were kept as audit trail. In other words, an audit trail for qualitative data collection consisting of raw data, data reduction and analysis was created and these were examined by other persons to confirm the findings.

Confirmability

Confirmability criterion concerns the degree of neutrality within a research the extent to which the findings of a study are shaped by the respondents and not researcher bias, motivation, or interest is based on the perspective that the integrity of findings lies in the data and the researcher must trace the conclusions back to the original sources (Lynch, 1996) by tying together the data, analytic processes, and findings clearly so that the reader is able to certify the adequacy of the findings (Morrow, 2005). Despite the acknowledgement that research is never objective, Morrow (2005) notes that confirmability addresses the core issue that "findings should represent, as far as is (humanly) possible, the situation being researched rather than the beliefs, pet theories, or biases of the researcher" (Gasson, 2004 in Morrow, 2005, p.252). Techniques to ensure confirmability entails confirmability audit, triangulation, detailed descriptions, reflections. As discussed above, triangulation of different data sources, thick description of each section and the field notes of the researcher are helpful to maintain confirmability.

3.8.1.2 Validity for the quantitative data

Assumptions of repeated measures ANOVA test were not violated. The specific assumptions included:

- The dependent variable, students' scores in different administrations of the test, is a continuous variable.

- The independent variable, students' gains in scores, consists of three categorical groups which correspond to their scores in reading, listening and writing.

- There were no significant outliers in the data sets.

- The distribution of the dependent variable was normally distributed in the related groups.

- Sphericity assumption was fulfilled. In other words, the variances of the differences between all combinations of related groups were equal.

A piloting study was conducted at the onset of the research study to ensure reliability of the data collection instruments.

3.8.2 Ethical issues

Certain measures were taken in order to protect the anonymity, privacy and confidentiality of the participants as well as the data. To begin with all participants took part in the study on a voluntary basis. They were provided with information about the nature and the purpose of the research as well as type of the data I needed, time frame they needed to invest to data collection process, how the data might be used and positive outcomes of the research. In addition, all participants were asked to fill in consent forms at the onset of their participation. Identity of all participants was concealed and collected data was kept out of reach of third parties.

3.9. Summary

The following table outlines the research questions of the study, how each research question is addressed and examined by different instruments and analysed.

Table 11

Outline of the research design

Research Questions	Participants	Data Collection Instrument(s)	Data Analysis	Purpose of the Analysis
1. Is there a potential washback effect of the TRACE exam on teaching (methods, materials, and tasks)?	Instructors and students	Class observation scheme Questionnaires Interviews	Descriptive statistics Lists of frequency counts of responses Pattern coding (Bogdan and Biklen, 1998)	To investigate ELT teachers' and students' perceptions as to how TRACE affects students' learning and teaching methodology
2. How does the language instruction program, based on EAP skills, result in gains in	Students and instructors	TRACE Exam that students take at the beginning of the academic year	Descriptive Statistics	To examine the difference in test scores between two

scores on the writing, listening, and reading parts of the TRACE exam?		and TRACE exam that upper and advanced level students take in January after instruction in the preparatory program	Correlations on the subsections of the TRACE Reliability pre-post for different administrations of TRACE ANOVA for repeated measures	administrations, which will serve as a measure of progress
3. How effective are the decisions made by TRACE in identifying the language competency required for academic study within this university program over the time when students embark on their further academic studies at their departments?	University instructors of mainstream courses Freshman students	Questionnaires Interviews	Descriptive statistics Pattern coding (Bogdan and Biklen, 1998)	To track an assumed link between the TRACE test results the use of English for Academic Purposes with respect to language and academic skills at mainstream departmental university courses.

CHAPTER 4

RESULTS

4. 1. Introduction

This chapter presents the results of the study regarding washback effect and instructional sensitivity of TRACE on teaching and learning. Viability of test decisions over time, in other words, consequential validity was also examined within the scope of this study. Data were gathered from pre- and post- proficiency exam scores, questionnaire about the student and teacher perceptions related to test effect and quality of learning, teacher and student questionnaire regarding consequential validity, semi-structured interviews, and real-life samples of Freshman students who passed on to their departments upon taking the proficiency exam-TRACE.

Data analysis was done to in three stages to answer the research questions asked on three major issues addressed in this study. The first set of analyses was on the data collected for washback effect. More specifically, the analysis was conducted to answer the research question of whether there was a potential washback effect of the TRACE exam on teaching (methods, materials, and tasks).

The second set of analysis was on the data collected for instructional sensitivity. In other words, the analysis attempted to answer second research question which focused on whether TRACE is sensitive to instruction. To this end, data analysis shed light on whether language instruction program, based on EAP skills, resulted in gains in scores on the writing, listening, and reading parts of the TRACE exam.

Finally, the third set of analysis focused on data gathered for the consequential validity of TRACE. Data analysis examined effectiveness of the decisions made by TRACE in

identifying the language competency required for academic study within this university program over the time.

4.2. Findings related to washback and instructional sensitivity

4.2.1 Findings of the student questionnaire for washback

English Language Background of the respondents

Overwhelming majority of the respondents, (94 %) did not live with a parent or a close relation who was a native speaker of English. The data revealed that 94% of the respondents were not exposed to English in their family when they were children, 39% of the respondents started studying English as a foreign language in kindergarten between the ages of 3 and 6, 41,5% received instruction in English at primary school and 19,5% were introduced to English at the secondary school. It is important to note that despite the fact that all respondents indicated that they were exposed to English in their formal schooling as off early ages before they started university, they could not receive high enough scores on TRACE to be exempted from the prep program at the beginning of the semester.

Attitudes to Teaching Materials, Tasks, and TRACE

In the second part of the questionnaire, initially students were asked to respond to 5 items surveying their opinions about TRACE in the form of 4 options-Likert type questions. Students' responses are outlined in the table below.

Table 12

Students' perceptions of TRACE

	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
1. I understand what English language (grammar and vocabulary presented in the prep program) and academic skills are required by TRACE.	3	2	7	6	79	64	33	27
2. I think the skills tested in TRACE are necessary to learn	8	7	21	17	67	55	26	21
3. I think TRACE tests what I will be asked to do at my department after prep school.	13	11	24	20	66	54	18	15
4. I feel TRACE is an appropriate test as an exit exam for prep school students.	9	7	32	26	57	46	23	19
5. I think TRACE measures my language ability level and academic skills accurately.	6	5	35	29	58	47	22	18

It has been construed that students' learning process and product can be influenced by their attitudes toward and perceptions of the components and purpose of a test they take (Hughes, 1993). Therefore, a positive attitude toward a certain test and awareness of the nature of the test are considered important for positive washback effect (Weili, 2010). The findings indicated that 91% of the respondents claimed to have an awareness of academic and language skills required by the TRACE. Interestingly, 71% of the respondents claimed that skills tested by TRACE were 'necessary to learn'. On the other hand, for 24% of the respondents, skills tested by TRACE were not necessary to learn. Also 31% of the students stated a discrepancy between tested skills on TRACE and skills that they would be required to master later on in their departments. In contrast, for 69% of the respondents, TRACE had ecological validity because they thought there was a correspondence between the language requirements of the undergraduate schools and the instructional materials offered in PEP. As for 'appropriateness of the test as an exit exam for PEP, student response displayed variety. The results indicated that 65% of the students disclosed positive perceptions whereas 33 % stated negative perceptions about TRACE and 91% of the respondents claimed that they were knowledgeable about the content of the exam in terms of tested skills, tasks and exam format.

Findings related to teaching

Student perceptions of how often certain tasks were employed by their teachers are outlined in the table below.

Table 13

Frequency and percentages for students' perceptions of tasks in the classroom

Frequency of tasks in the English classes.	Never		Sometimes		Frequently		Always	
	f	%	f	%	f	%	f	%
1. Discussions with a partner	3	2	40	33	53	43	27	22
2. Discussions in small groups	3	2	49	40	47	38	24	20
3. Writing essays	4	3	43	35	47	38	29	24
4. Self and/or peer evaluation of essays using criteria given by the teacher	5	4	35	29	50	41	32	26
5. Games, puzzles, quizzes	7	6	57	46	37	30	22	18
6. Group interviews, and other project work	12	10	61	50	39	32	10	8
7. Creative writing	24	20	49	40	37	30	12	10
8. Web/on-line activities	4	3	38	31	49	40	32	26
9. Role plays	47	38	37	30	28	23	10	8
10. Grammar exercises	5	4	20	16	54	44	44	36
11. Skills training (how to read, listen, speak, learn vocabulary and write better)	3	2	22	18	56	46	42	34
12. Vocabulary exercises	0	0	23	19	46	37	54	44
13. Exam practice	2	2	44	36	48	39	28	23
14. Listening to lectures and taking notes to answer questions	4	3	27	22	59	48	33	27
15. Reading a text and answering questions	1	1	24	20	57	46	41	33
16. Reading texts from books and other materials to use information in writing an essay	9	7	42	34	49	40	23	19
17. Reading different texts and answering questions that compare these different reading texts	6	5	44	36	47	38	24	20
18. Using notes (from reading or listening) to respond to a writing question in essay form	5	4	39	32	50	41	28	23
19. Working on skills and strategies about how to take a test	13	11	41	33	42	34	27	22

Data analysis suggests that 65% of the respondents claimed discussions with a partner was a frequent classroom task. Also, 58% suggested discussion in small group was a recurrent

activity. Additionally, for 66% of the respondents, web-online activities were often employed in the classroom. For 50% of the respondents, group interviews and other project work were tasks that were sometimes done in the class. Furthermore, students' response revealed that except for role-play, they concurred that a range of tasks, such as games, puzzles, quizzes, sometimes (%46) and frequently (48%) took place in the classrooms. Similarly, 40% of the respondents held the belief that creative writing was a frequent classroom task. As for using ideas from various sources into one's own work, students claimed that teachers sometimes employed tasks such as "reading texts from books and other materials to use information in writing an essay" (34%), "reading different texts and answering questions that compare these different reading texts" (36%) and "using notes (from reading or listening) to respond to a writing question in essay form" (32%). It is important to note here that these tasks, which encourage discourse synthesis, were reported to be frequently and always used by a large percentage of the students, mostly indicating frequent exposure. As for tasks which encouraged reflective thinking, students' response indicate that self and/or peer evaluation of essays was sometimes (29%) and often (71%) exercised. Finally, students stated that tasks related to exam practice (62%) and exam taking skills and strategies (56%) were frequently and always done.

Data from student questionnaire about classroom tasks employed by their teachers indicated that there was a balance of instruction of four skills, that there was a focus on synthesizing ideas across different sources, and that there were attempts to exploit communicative activities which foster student to student interactional patterns. Based on students' account, the fact that students would take TRACE at the end of instruction did not lead to gearing the curriculum towards the test content. For example, it was claimed that working on speaking was not abandoned even though it is not tested in the proficiency exam. However, data revealed that exam practice and exam taking skills and strategies took up class time on a large scale.

Findings related to course materials

The questionnaire items regarding students' perceptions of materials catering for reading skills generally revealed positive opinions. A summary of these opinions is given in the table below.

Table 14

Frequency and percentages for students' perceptions of materials catering for Reading Skills

Materials Covering Reading Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
1.Our materials covered reading skills such as the following that helped me to learn English well:								
1.1 understanding main ideas	2	2	13	11	76	62	32	26
1.2 understanding the writer's overall point,	1	1	15	12	79	64	28	23
1.3 inferring purpose, tone, and attitude of the writer	0	0	17	14	80	65	26	21
1.4 guessing meanings of unknown words	3	2	20	16	78	63	22	18
1.5 analysing organisation of the text	4	3	16	13	81	66	22	18
1.6 synthesizing information from different texts to answer questions	5	4	36	29	66	54	14	11
2. Our materials covered reading skills that would help me to perform well on the test.	5	4	25	20	75	61	17	14
3.There was a variety of reading text types (short texts and long texts)	3	2	17	14	79	64	21	17
4. Reading texts were followed by different activities such as:								
4.1 multiple choice,	0	0	14	11	63	51	46	38
4.2 summarising	5	4	25	20	63	51	30	24
4.3 paraphrasing	3	2	12	10	83	68	25	20
4.4 matching	0	0	22	18	69	56	31	25
4.5 open-ended questions	3	2	24	20	66	54	30	24
4.6 fill-in blanks	1	1	16	13	76	62	30	24
4.7 choose the sentence that completes a paragraph	1	1	18	15	76	62	28	23
4.8 write or choose a title for the text	5	4	15	12	73	59	30	24

The respondents generally claimed positive perceptions about materials in nurturing certain reading skills. 88% of the responses claimed that materials covered understanding the main ideas. In addition, for 86% of the students, materials covered exercises that helped them to read for writer's overall point, inferring purpose, tone and attitude. Similarly, 81% of the

students claimed that reading materials catered for guessing vocabulary. 84% of the students suggested that reading materials also covered tasks that tapped on to analysing organisation of the text. While 65% concurred that the materials involved synthesising information skill. Surprisingly, 33% of the respondents claimed that reading materials did not cater for synthesising information from different texts to answer questions. In addition, as for the relation between reading skills coverage of the materials and students' performance on the test, except for 76% of the respondents indicated positive perceptions. It was also generally agreed that reading texts were followed by different activities. However, 22% of the students claimed that reading texts excluded open-ended questions and summarising exercises.

The survey questions about coverage of the course materials with respect to listening skills and how much materials contributed to students' learning English elicited mostly positive response from a large majority. Students' response to how well they thought materials nurtured improvement of their listening skill are summarised below.

Table 15

Frequency and percentages for students' perceptions of materials catering for listening skills

Materials Covering Listening Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
5. Our materials covered listening skills such as the following that helped me learn English well								
5.1 understanding overall point	3	2	14	11	86	70	20	16
5.2 inferring purpose, tone, and attitude of the speaker(s)	4	3	14	11	82	67	23	19
5.3 guessing meanings of unknown words	6	5	26	21	74	60	17	14
5.4 noting down information and using these notes to answer questions	8	7	20	16	67	55	28	23
5.5 analysing the organisation of the listening text to understand main and supporting ideas	5	4	20	16	79	64	19	15
5.6 combining my listening notes with other sources to write an essay.	6	5	25	20	69	56	23	19
6. Our materials covered listening skills that would help me perform well on the test.	6	5	23	19	72	59	21	17

7. There was a variety of listening text types (e.g. conversations, talks, TV shows and lectures)	4	3	10	8	71	58	36	29
8. Listening texts were followed by different activities such as:								
8.1 multiple choice,	3	2	7	6	70	57	43	35
8.2 summarising	10	8	31	25	58	47	24	20
8.3 matching	6	5	17	14	70	57	30	24
8.4 open-ended questions	8	7	25	20	64	52	26	21
8.5 fill-in the blanks	3	2	17	14	73	60	30	24
8.6 paraphrasing	7	6	26	21	66	54	23	19
8.7 while listening	5	4	10	8	78	63	29	24
8.8 lecture listening	2	2	8	7	76	62	37	30

Students claimed that the listening materials covered a variety of skills including listening for main ideas (86%), inference (86%), guessing vocabulary form context (74%), note-taking (78%), analysing the organisation of on oral text (79%) and combining notes with other sources (75%) on a large scale and this assisted their learning English. However, it should be noted 25% of the respondents disagreed that materials covered certain skills such as 'combining listening notes with other sources to write an essay' which relates to an important construct in TRACE; discourse synthesis. Additionally, 26% of the students remarked the exclusion of 'guessing meanings of unknown words'. Also 20% of the student response claimed that listening materials covered analysing organisation. In addition, 24% of the students held the belief that listening materials would not help them perform well on the test. Surprisingly, 76% indicated that listening materials would contribute positively to their exam performance. Besides, majority agreed that there was a variety of listening texts and these were accompanied by different activities such as multiple choice follow up questions (92%), lecture listening (92%), while listening (87), fill in the blanks (84%), and matching (81%). As for summarising, paraphrasing and responding to open-ended questions there were some students who reported negative perceptions. Although summarising and paraphrasing skills are not tested explicitly in TRACE they constitute an important component of discourse synthesis which requires students to use ideas across various sources. 33% of the students indicated that summary activities were

missing and 26% expressed exclusion of paraphrasing activities in materials. Also, 27% of the students stated that listening materials did not have open-ended questions.

The following table outlines students' perceptions of materials catering for writing skills.

Table 16

Frequency and percentages for students' perceptions of materials catering for writing skills

Materials Covering Writing Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
9. Our materials covered writing skills such as the following that helped me learn English well								
9.1 organising ideas from my notes into a written plan	2	2	17	14	78	63	26	21
9.2 introducing and narrowing down ideas	4	3	18	14	83	68	17	14
9.3 organising my ideas in a logical way	2	2	13	11	85	69	23	19
9.4 using information and arguments from different sources into my writing	2	2	22	18	75	61	21	17
9.5 evaluating my own work and revising	2	2	22	18	77	63	22	18
9.6 evaluating my peers' writing based on criteria given by the teacher	2	2	20	16	72	59	28	23
10.Our materials covered writing skills that would help me perform well on TRACE.	5	4	17	14	69	56	31	25
There was a variety of opportunities for writing including:								
11.1 writing reviews	9	7	28	23	61	50	25	20
11.2 creative writing	8	7	28	23	61	50	26	21
11.3 report writing	6	5	35	29	64	52	18	15
11.4 writing a response to a text	10	8	25	20	70	57	18	15
11.5 story writing	12	10	40	33	51	42	19	15
11.6 summary writing	5	4	24	20	72	59	22	18
12. I made use of information given in reading and listening texts in the writing activities.	3	2	20	16	66	54	31	25

Data analysis elucidated that 84% of the students agreed that course materials covered writing skills such as organising notes. Also for 82% of the students writing materials involved tasks for introducing and narrowing down ideas. Similarly, 88% of the students claimed that writing materials covered organising ideas. Also 81% of the respondents suggested that materials helped the respondents to improve their competency through self and peer evaluation.

It can be inferred that these positive perceptions indicate in-house prepared supplementary materials which guided students during brainstorming and outlining phases of process writing component of the preparatory program since course books at the levels did not have any writing input. Similarly, positive washback effect on materials was suggested not only in student but also teacher interviews. Although 61% of the students agreed and 17% strongly agreed that materials covered the skill of using information and arguments from different sources into writing, 20% of the respondents disagreed that materials guided students for discourse synthesis. This statement received the highest percentage of negative student perception among all.

In addition, the students held the belief that materials covered writing skills that helped them boost their performance in the proficiency test. However, response signified that some students thought materials did not provide a variety of opportunities in story writing (43%), report writing (34%), writing reviews (30%), creative writing (30%), writing a response (28%), and summary writing (24%). This can also be regarded an indication of close alignment between materials and the tested skills in TRACE and this response was noteworthy because generally students agreed with the statements on the questionnaire but conveyed negative perceptions regarding scarcity of the above mentioned writing tasks. Finally, a large percentage of the students stated that the made use of information given across different sources in their written outcome.

Although speaking is not one of the skills tested in TRACE and highly positive student perceptions were revealed regarding how well the teaching materials covered speaking skills. For the overwhelming majority of the students, course materials covered speaking skills ranging from making polite requests (84%), starting and ending discussions (85%), explaining ideas (87%), giving examples (87%), (dis) agreeing and asking (91%) and asking for clarification (87%). Furthermore, most students concurred there were a range of the opportunities and activities provided by the materials. However, some students noted the scarcity of some certain

activities including role-play (37%), story-telling (%36) and making use of information across different sources in spoken production. In contrast other activities which received higher percentages of agreement, these do not represent the task types that were graded throughout upper- intermediate and advanced level as a scope of the portfolio and performance assessment component (i.e. giving presentations and participating in group discussions).

Table 17

Frequency and percentages for students' perceptions of materials catering for speaking skills

Materials Covering Speaking Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
13. Our materials covered speaking skills such as the following that helped me learn English well:								
13.1 making polite requests	4	3	15	12	82	67	21	17
13.2 starting and ending discussions in an appropriate way	1	1	17	14	79	64	26	21
13.3 explaining my ideas	1	1	15	12	78	63	29	24
13.4 giving examples	2	2	13	11	79	64	28	23
13.5 agreeing and disagreeing	1	1	10	8	77	63	34	28
13.6 asking for clarification questions (e.g. Could you explain your words?)	3	2	11	9	77	63	29	24
14. There was a variety of opportunities and activities for speaking including:								
14.1 pair work discussions	2	2	18	15	74	60	29	24
14.2 role-play	14	11	32	26	55	45	22	18
14.3 group work discussions	5	4	13	11	73	59	31	25
14.4 presentation	3	2	19	15	69	56	32	26
14.5 asking and answering questions	1	1	16	13	76	62	30	24
14.6 story telling	13	11	31	25	55	45	24	20
15. I made use of information given in reading and listening texts in the speaking activities.	5	4	29	24	57	46	29	24

Finally, when the students were asked about grammar and vocabulary materials most positive perceptions were revealed.

Table 18

Frequency and percentages for students' perceptions of materials catering for grammar and vocabulary

Materials Covering Grammar and Vocabulary	Never		Sometimes		Frequently		Always	
	f	%	f	%	F	%	f	%
16. Our materials covered grammar & vocabulary that are important for learning English	1	1	22	18	57	46	43	35
17. Our materials covered grammar & vocabulary that are important to do well on tests.	5	4	23	19	48	39	46	37
18. There was a variety of grammar & vocabulary activities	2	2	25	20	54	44	42	34
19. (Grammar & vocabulary lessons were followed by different activities such as:								
19.1 multiple choice	2	2	9	7	72	59	40	33
19.2 matching	0	0	15	12	69	56	39	32
19.3 fill-in the blanks	1	1	17	14	65	53	40	33
19.4 using newly learned grammar and vocabulary in writing	3	2	13	11	65	53	42	34
19.5 using newly learned grammar and vocabulary in speaking	7	6	12	10	61	50	43	35

According to 81% of participating students upper-intermediate and advanced level course materials covered grammar and vocabulary that are important for improving their language competency. Also 76% of the respondents claimed that these materials helped them improve their exam scores (76%). 78% of the respondents agreed that there was variety of grammar and vocabulary activities and the lessons were accompanied by a range of activities.

Student perceptions regarding materials and how well they nurtured students' improvement in language skills captured in the data above suggested that teaching materials and TRACE were closely aligned and this alignment corresponded to the TRACE's positive washback on materials. Students deemed materials covering tested skills and echoing exam tasks and questions types as a positive contribution to their learning.

4.2.2. Findings of the student questionnaire for instructional sensitivity

When students were surveyed related to how much they learned in reading skills as a result of PEP majority of the respondents conveyed positive perceptions.

Table 19

Students' perceptions of how much they learned in terms of reading skills

.	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
1. Find out the main idea(s) of the text.	2	2	9	7	83	68	28	23
2. Identify supporting ideas, details and examples.	1	1	10	8	81	66	31	25
3. Find out the message of the text by looking at visual and contextual clues in the text.	1	1	16	13	74	60	31	25
4. Identify specific information by using some strategies (such as reading carefully and taking reading notes).	3	2	11	9	79	64	30	24
5. Relate ideas in a reading to one another by understanding how the text is organised, whether ideas follow one another, whether the writer explains his ideas by explanations and examples.	2	2	12	10	87	71	21	17
6. Guess the meaning of unknown words using my knowledge.	5	4	21	17	71	58	25	20
7. Read between the lines to infer the tone and purpose.	1	1	17	14	74	60	30	24
8. Understand words and phrases such as "This shows …, These are…" that refer to objects, people or concepts.	2	2	15	12	78	63	28	23
9. Recognize different organisations of ideas in the reading text to understand the author's overall point of view.	0	0	15	12	82	67	26	21
10. Select and make notes from reading texts to answer questions.	3	2	17	14	77	63	25	20
11. Combine my reading notes with other sources to write an essay.	2	2	19	15	70	57	32	26
12. Compare and contrast different texts to understand why they are written (purpose), have different styles (tone, author's attitude), for whom the text is written (audience)	3	2	20	16	74	60	25	20

The findings indicate that the students considered they had progressed on a large scale with respect to reading skills. Their response revealed a large proportion of agreement to all of the statements which required a self-evaluation of the listed reading skills. To illustrate, 91% of the students agreed that they learned how to find the main ideas and identify supporting ideas, details and examples. Even for the cognitively challenging reading skills, such as comparing and contrasting different texts to infer the author's attitude, purpose, tone and audience, there was 80% of student agreement that they learned.

Similarly, a very large proportion of the students deemed that they learned a variety of listening skills such as understanding the relationship between the ideas (88%), using notes to answer questions (84%), follow the arguments within the text (82%), listening for the main

ideas (82%), identifying specific information (82%), inferring speaker's attitude and purpose (81%), understanding use of stress and intonation (78%), and recognizing organisation in the text (77%). However, students' perception regarding guessing new words and combining listening notes do not augur well with the high percentages of student agreement with other listening skills since 30% disagreed with the former and 24% disagreed with the latter.

Table 20

Students' perceptions of how much they learned in terms of listening skills

.	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
1. Guess the meaning of new words to understand a listening text	8	7	26	21	66	54	22	18
2. Find out the main ideas based on the clues in the listening text such as organisation of the listening text, examples and key words used and questions asked by the speaker(s).	3	2	20	16	71	58	29	24
3. Identify specific information (by paying attention to names, dates, numbers, signal words such as one of the important problems…, firstly,…etc.)	6	5	16	13	75	61	26	21
4. Understand the relationship between ideas	3	2	13	11	83	68	24	20
5. Understand the speaker's use of stress and intonation to signal different sections of a listening text.	4	3	23	19	71	58	24	20
6. Recognize organisation of different sections of a listening text.	6	5	21	17	71	58	23	19
7. Select and take notes from an organised listening text.	4	3	20	16	73	59	26	21
8. Use my listening notes to answer questions.	5	4	14	11	73	59	31	25
9. Combine my listening notes with other sources to write an essay.	5	4	25	20	68	55	24	20
10. Infer who speakers are, what the situation is, speaker's attitude and purpose	4	3	19	15	75	61	25	20
11. Follow the arguments of talks and presentations if they are organised and on a familiar topic.	5	4	17	14	80	65	21	17

As for how much students learned with respect to writing skills, data analysis revealed large majority of the students held the belief that they improved their writing skills to a large extent in all sub-writing skills. However, for certain skills such as taking organised notes while reading (28%), supporting ideas using explanations, descriptions, statistical data, information

from a variety of sources (22%), choosing appropriate organisational patterns (21%), summarising or paraphrasing information from other sources (24%) and editing essay based on feedback from peers (22%) nearly one fourth of the response reported disagreement. Students' response is outlined in the table below.

Table 20

Students' perceptions of how much they learned in terms of writing skills

	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
1. Give short answers to the questions about the texts I read or listen to	6	5	13	11	72	59	32	26
2. Complete missing info about the texts I read or listen to	4	3	18	15	72	59	29	24
3. Take organised notes while listening	3	2	22	18	73	59	25	20
4. Take organised notes while reading	2	2	32	26	63	51	25	20
5. Organise my ideas in an essay, generate content, link my ideas to one another in paragraphs and in the whole essay by explaining, using linking words (e.g. In addition, on the other hand, however, furthermore…etc.) and giving examples.	1	1	20	16	70	57	32	26
6. Generate content, organise my ideas in to a plan/an outline, introduce the topic and relate my ideas to each other by using linking words	1	1	23	19	66	54	33	27
7. Support my ideas using explanations, descriptions, statistical data, information from other sources and examples	2	2	24	20	67	55	29	24
8. Conclude my essay by summarising main points and offering my final comments on the topic.	2	2	19	15	70	57	31	25
9. Choose appropriate organisational patterns (such as defining concepts, comparing and contrasting, describing cause and effects …etc.) to develop my ideas in the paragraphs	2	2	23	19	71	58	27	22
10. Write essays of about 400-450 words in which I can organize and express my ideas well and use information from other sources.	2	2	21	17	63	51	36	29
11. Summarise or paraphrase information from other sources such as articles and lectures	2	2	27	22	67	55	26	21
12. Re-write my essay by getting feedback from my class mates	5	4	22	18	64	52	32	26
13. Re-write my essay by getting feedback from my teacher	0	0	15	12	67	55	41	33
14. Edit my own work by self-evaluating it	1	1	17	14	72	59	32	26
15. Write in an academic style by using mostly correct grammar, spelling and punctuation, as well as	1	1	20	16	68	55	30	24

variety of grammar structures and appropriate
vocabulary.

The following table presents the data analysis regarding students' opinions of progress in their language competency.

Table 22

Students' perceptions of how much they learned in terms of grammar and vocabulary

.	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
1. Record and remember newly learned vocabulary items in an organised way	6	5	13	11	76	62	27	22
2. Use newly learned vocabulary in my speaking	2	2	14	12	82	67	25	20
3. Use newly learned vocabulary in my writing	4	3	11	9	66	54	42	34
4. Use dictionaries effectively	7	6	17	14	64	52	34	28
5. Use different grammar structures accurately	3	2	15	12	68	55	36	29

Overwhelming majority of students reported positive perceptions regarding their progress in grammar and vocabulary stating that they learned how to record and remember vocabulary (84%), use newly learned lexis while they speak (87%) and write (88%), use dictionaries effectively (80%) and use a range of grammar structures accurately (84%).

4.2.3 Findings of the teacher questionnaire for washback

21 PEP instructors teaching at upper-intermediate and advanced levels took the questionnaire which was designed to survey their perceptions related to washback and instructional sensitivity of the TRACE exam. The first part of the questionnaire surveyed 'Attitudes to Teaching Materials, Tasks and TRACE' and this subscale consisted of 91 items ($\alpha = .95$). The final part of the questionnaire surveyed participants' perceptions regarding how

much learning happened (instructional sensitivity) and Cronbach's alpha for 43 items was α =.61. The questionnaire was found to be highly reliable (134 items; α = .91).

In Part 1 of the questionnaire, teachers were surveyed regarding their ideas about TRACE. They indicated their level of agreement to 5 items in the form of 4 option Likert type statements in Section A. Teachers' responses are summarised in the table below.

Table 23

Teachers' perceptions of TRACE

	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
I understand what English language (grammar and vocabulary presented in the prep program) and academic skills are required by TRACE.	0	0	0	0	9	43	12	57
I think the skills tested in TRACE are necessary to learn	0	0	0	0	14	67	7	33
I think TRACE tests what students will be asked to do at their department after prep school	1	5	5	24	12	57	2	10
I feel TRACE is an appropriate test as an exit exam for prep school students.	0	0	1	5	17	81	3	14
I think TRACE measures my students' language ability level and academic skills accurately.	0	0	4	20	13	62	4	19

Remarking an awareness of content of TRACE, all teachers agreed that they acknowledge the necessary skills and language required by this proficiency exam. In addition, all teachers concurred that they deemed all the tested skills as 'necessary to learn'. Teachers' response displayed a difference in comparison to student perceptions because 24% of the students disagreed that the tested skills were important to learn. Similarly, there was a discrepancy between students and teachers since except for one respondent all teachers held the opinion that TRACE was an appropriate exit exam for the students. However, 33% of student response indicated disagreement. Also, some teachers disagreed that TRACE tests what students will be asked to do at their department after school. In other words, 29% of the response revealed teachers' disbelief in scope of TRACE in testing authentic language use at the

departments. Overwhelming majority of the teachers (80%) believed that TRACE measures students' language ability and skills accurately. It can be inferred that teachers had more faith in the test with respect to accurate measurement in comparison to students because 34% of the students had stated negative perceptions regarding how well TRACE measures their performance.

Finding related to Washback on Teaching

When teachers were surveyed about their opinions of tasks they employed in the classroom majority of the teachers indicated that a range of communicative activities including having discussions with a partner (62%), discussions in groups (72%) were frequent activities and they sometimes made use of role-plays (57%), web-online activities (52%), group interviews and other project work (43%).

Table 24

Frequency and percentages for teachers' perceptions of tasks used in the classroom

Frequency of tasks in the English classes.	Never		Sometimes		Frequently		Always	
	f	%	f	%	f	%	f	%
1. Discussions with a partner	1	5	7	33	7	33	6	29
2. Discussions in small groups	0	0	6	29	10	48	5	24
3. Writing essays	0	0	7	33	9	43	5	24
4. Self and/or peer evaluation of essays using criteria given by the teacher	2	10	11	52	8	38	0	0
5. Games, puzzles, quizzes	0	0	9	43	9	43	3	14
6. Group interviews, and other project work	3	14	9	43	6	29	3	14
7. Creative writing	10	48	10	48	1	5	0	0
8. Web/on-line activities	0	0	11	52	7	33	3	14
9. Role plays	7	33	12	57	2	10	0	0
10. Grammar exercises	1	5	2	10	9	43	9	43
11. Skills training (how to read, listen, speak, learn vocabulary and write better)	0	0	3	14	10	48	8	38
12. Vocabulary exercises	0	0	3	14	9	43	9	43
13. Exam practice	2	10	7	33	9	43	3	14
14. Listening to lectures and taking notes to answer questions	0	0	2	10	13	62	6	29
15. Reading a text and answering questions	0	0	1	5	12	57	8	38
16. Reading texts from books and other materials to use information in writing an essay	0	0	7	33	8	38	6	29

17. Reading different texts and answering questions that compare these different reading texts	2	10	10	48	5	24	4	19
18. Using notes (from reading or listening) to respond to a writing question in essay form	2	10	10	48	7	33	2	10
19. Working on skills and strategies about how to take a test	1	5	8	38	8	38	3	14

Activities with a focus on exercising the tested skills such as writing essays (67%), grammar exercises (86%), vocabulary activities (86%), listening to lectures and taking notes to answer questions (91%), reading a text and answering questions (96%) were also frequently utilized. Activities which catered for critical thinking skills and autonomous learning revealed a variety of response. Skills training was reported to be done frequently in class by an overwhelming majority (86%) but 52% of the teachers reported that self and/or peer evaluation of essays was not a regular component of the lessons because it was at times done. However, another bedrock attribute of TRACE, discourse synthesis wasn't deemed to be used very often. 48% of the response denoted teachers sometimes worked on reading different texts and answering questions to compare these different sources and using reading and/or listening notes to respond to a writing question in essay form. Based on 67% of response, it seems that teachers broadly guided students to read a variety of sources and integrate information presented in these texts while students were writing their essays. As for coaching towards TRACE, teachers stated that they frequently (52%) and sometimes (38%) worked on skills and strategies about how to take a test. Similarly, exam practice was a frequent activity according to 57% of the respondents.

Findings of student perceptions and teacher perceptions of tasks used in the classroom displayed similarity almost in all aspects. To illustrate, both students (62%) and teachers (57%) concurred exam practice was a frequent component of instruction and not only students (59%) but also teachers (67%) affirmed that reading texts from books and other materials to use information in writing an essay was frequently exercised in class. However, there was diversity among teachers and students regarding self and peer evaluation of essays using a given criteria

and creative tasks. As for self and/or peer evaluation tasks students reported that they were frequently required to conduct these tasks where as teachers stated they used these tasks not as frequently as students indicated. Similarly, students expressed that creative writing tasks were at times made use of whereas nearly half of the teachers (48%) stated that they never use creative writing tasks. It can be stated that teachers perceived that they made more frequent use of the tasks that resembled exam tasks which would prepare students and help them to increase their exam performance.

Student and teacher questionnaire data implied that although exam practice was integrated into teaching and changed what they taught, there wasn't any negative washback of TRACE regarding how teachers taught because both parties agreed that there was practice of 4 skills through some creative and communicative activities. However, there were disagreement regarding use of some collaborative and communicative activities which called for critical thinking and using content and ideas across different sources. Additionally, not only students but also teachers indicated that curriculum was not narrowed down to tested skills, excluding practice on speaking which was not assessed on TRACE.

Findings related to washback on materials

Teacher questionnaire on washback examined perceptions towards how well the materials catered for skills and sub-skills, whether materials were believed to assist students regarding gains in scores on the test and scope of activities that accompanied these materials. The following table outlines the opinions of the teachers regarding how well the materials cater for improvement of reading skills.

Table 25

Frequency and percentages for teacher perceptions of materials catering for reading skills

Materials Covering Reading Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
1. Our materials covered reading skills such as the following that helped my students to learn English well:								
1.1 understanding main ideas	0	0	0	0	12	57	9	43
1.4 understanding the writer's overall point,	0	0	2	10	11	52	8	38
1.5 inferring purpose, tone, and attitude of the writer	0	0	0	0	15	71	6	29
1.4 guessing meanings of unknown words	0	0	1	5	14	67	6	29
1.5 analysing organisation of the text	1	5	5	24	11	52	4	19
1.6 synthesizing information from different texts to answer questions	0	0	8	38	11	52	2	10
2. Our materials covered reading skills that would help my students to perform well on the test.	1	5	2	10	13	62	5	24
3. There was a variety of reading text types (short texts and long texts)	0	0	8	38	12	57	1	5
4. Reading texts were followed by different activities:								
4.1 multiple choice,	0	0	0	0	12	57	9	43
4.2 summarising	2	10	9	43	8	38	2	10
4.3 paraphrasing	2	10	10	48	8	38	1	5
4.4 matching	0	0	2	10	14	67	5	24
4.5 open-ended questions	0	0	4	20	15	71	2	10
4.6 fill-in blanks	0	0	4	19	14	67	3	14
4.7 choose the sentence that completes a paragraph	1	5	3	14	15	71	2	10
4.8 write or choose a title for the text	1	5	4	19	13	62	3	14

Findings pointed out that except for skills that tap on discourse synthesis and critical thinking nearly all teachers agreed that materials covered skills such as reading for main ideas, inferring and guessing meanings of unknown words. This finding may point out to a discrepancy between the positive washback intended by the test designers and how integration of skills is operationalised and reinforced through reading materials. TRACE is integrated and theme-based and therefore it necessitates summarising, paraphrasing, analysing organisation of a text to find main ideas and discourse synthesis. Some teachers reported that these skills were excluded from the materials. On the other hand, positive perceptions regarding other statements show that supplementary materials, which were designed specifically according to the test specifications, were regarded highly positively. However, 38 % of the response indicated lack

of materials which steered students into synthesizing information from different texts to answer questions and 29% marked exclusion of analysing organisation of a text within materials. Also, 38 % of the teachers deemed there weren't a variety of types of reading texts. It was stated that materials lacked summarising (53%), paraphrasing (58%) and open-ended questions (20%). It is important to note that summarising and paraphrasing are of vital importance for integrating ideas across sources into one's own work. However, they are not directly tested on the TRACE and they are not practiced frequently though activities in class and teaching materials. Therefore, I would like to note that although teachers denoted their overall positive perceptions about reading materials, some also pointed out that materials did not assist students adequately in skills related to citing sources.

The following table summarizes teacher perceptions regarding listening materials.

Table 26

Frequency and percentages for teachers' perceptions of materials catering for listening skills

Materials Covering Listening Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
5.Our materials covered listening skills such as the following that helped my students learn English well								
5.1 understanding overall point	0	0	1	5	13	62	6	29
5.2 inferring purpose, tone, and attitude of the speaker(s)	0	0	1	5	15	71	5	24
5.3 guessing meanings of unknown words	0	0	4	20	14	67	3	14
5.4 noting down information and using these notes to answer questions	0	0	0	0	14	67	7	33
5.5 analysing the organisation of the listening text to understand main and supporting ideas	2	10	3	14	12	57	4	20
5.6 combining their listening notes with other sources to write an essay.	1	5	9	43	9	43	2	10
6.Our materials covered listening skills that would help my students perform well on the test.	0	0	4	20	13	62	3	14
7.There was a variety of listening text types (e.g. conversations, talks, TV shows and lectures)	2	10	10	48	5	24	3	14
8.Listening texts were followed by different activities such as:								
8.1 multiple choice,	0	0	0	0	10	48	11	52
8.2 summarising	3	14	10	48	7	33	1	5
8.3 matching	1	5	6	29	13	62	1	5
8.4 open-ended questions	0	0	5	24	13	62	3	14

8.5 fill-in the blanks	0	0	4	19	13	62	4	19
8.6 paraphrasing	3	14	10	48	8	38	0	0
8.7 while listening	0	0	3	14	12	57	6	29
8.8 lecture listening	0	0	0	0	11	52	10	48

Almost all teachers agreed that listening materials helped students understand the main ideas (91%), infer the purpose, tone, and attitude of the speakers (95%), and take notes and use these in answering follow up questions (100%). However, some teachers indicated that listening materials did not facilitate learning since materials did not focus on integrating listening notes with other sources to write an essay (48%). This finding resonates the student response (24%) which also pointed out scarcity of focus in listening materials towards utilizing listening notes in writing. Also, 24% of the teachers believed that materials did not cater for analysis of the organisation of the listening text to understand the main ideas. 20% of the students have also agreed with the teachers regarding this finding.

In addition, a large majority of the teachers stated that materials covered listening skills that would help students increase their grades on the proficiency test. It is important to note here that both teachers and students agreed that listening materials had a facilitating effect on students' performance. In terms of variety of listening texts 58% response disagreed that there was a variety ranging from conversations to talks. Finally, findings reveal that most of the teachers agreed listening texts were followed by different activities such as multiple choice (100%), lecture listening (100%), matching (67%), open-ended questions (76%), fill-in-the blanks (81%), and while listening (86%). However, a large proportion of the teachers concurred that materials did not involve summarising (62%) and paraphrasing (62%) activities.

Findings related to teacher perception of materials catering for writing skills of the students revealed generally positive perceptions from a large majority of the respondents. Overwhelming majority of the teachers enunciated that materials covered an array of skills such as organising ideas from notes into an outline (90%), introducing and narrowing down ideas

(81%), organizing ideas in a logical way (91%) and self-evaluation of one's own work (77%).

On the other hand, it was the contention of some participants that writing materials did not cater for certain skills including using information and arguments from different sources into writing (48%) and revising one's own work based on peer evaluation (39%). In upper- intermediate and advanced levels writing materials comprised of in-house prepared supplementary materials and although great majority expressed highly positive perceptions about these materials the above mentioned skills deemed to be lacking. It is important to note that the former of these skills represents an important underlying feature of TRACE; discourse synthesis and the latter signifies one of the life skills of PEP curriculum; autonomous learning. Not only some teachers but also some students expressed that materials did not tap on to these skills. Teacher perceptions of writing materials are outlined below.

Table 27

Frequency and percentages for teachers' perceptions of materials catering for writing skills

Materials Covering Writing Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
9.Our materials covered writing skills such as the following that helped my students learn English well								
9.1 organising ideas from their notes into a written plan	0	0	2	10	10	57	7	33
9.2 introducing and narrowing down ideas	0	0	4	20	14	67	3	14
9.3 organising their ideas in a logical way	0	0	1	5	15	71	5	24
9.4 using information and arguments from different sources into their writing	2	10	8	38	11	52	0	0
9.5 evaluating their own work and revising	1	5	4	19	14	67	2	10
9.6 evaluating peers' writing based on criteria given by the teacher	2	10	6	29	13	62	0	0
10.Our materials covered writing skills that would help my students perform well on TRACE.	1	5	2	10	15	71	3	14
11.There was a variety of opportunities for writing including:								
11.1 writing reviews	4	20	12	57	4	20	1	5
11.2 creative writing	8	38	9	43	4	20	0	0
11.3 report writing	8	38	11	52	2	10	0	0
11.4 writing a response to a text	3	14	10	48	8	38	0	0
11.5 story writing	7	33	12	57	2	10	0	0
11.6 summary writing	1	5	4	19	15	71	1	5
12. Students made use of information from reading and listening texts in the writing activities.	0	0	4	19	12	57	1	5

Overwhelming majority of the teachers affirmed that materials had a coverage which aimed at improving speaking skills such as 'explaining ideas' (91%), 'starting and ending discussions' (91%), 'agreeing and disagreeing' (95%), 'giving examples' and asking for clarification questions (81%). Among speaking skills there was disagreement about only "making polite request" since 43% of the response was negative. Overall teachers conveyed positive perceptions about the materials and stated that the TRACE had a positive influence on the materials.

Table 28

Frequency and percentages for teachers' perceptions of materials catering for speaking skills

Materials Covering Speaking Skills	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
13. Our materials covered speaking skills such as the following that helped their students learn English well:								
13.1 making polite requests	3	14	6	29	9	43	2	10
13.2 starting and ending discussions in an appropriate way	2	10	0	0	16	76	3	15
13.3 explaining their ideas	1	5	1	5	15	71	4	20
13.4 giving examples	1	5	1	5	16	76	3	14
13.5 agreeing and disagreeing	1	5	0	0	16	76	4	19
13.6 asking for clarification questions (e.g. Could you explain your words?)	1	5	3	14	13	62	4	19
14. There was a variety of opportunities and activities for speaking including:								
14.1 pair work discussions	2	10	2	10	12	57	5	24
14.2 role-play	4	19	12	57	5	24	0	0
14.3 group work discussions	0	0	1	5	14	67	6	29
14.4 presentation	0	0	2	10	12	57	7	33
14.5 asking and answering questions	0	0	1	5	14	67	6	29
14.6 story telling	8	38	9	43	4	20	0	0
15. My students made use of information given in reading and listening texts in the speaking activities.	1	5	5	24	13	62	1	5

In addition, except for certain activities such as role-play and story-telling majority of the teachers agreed that there was a variety of opportunities and activities for speaking within the course materials. Nearly all teachers stated that there were abundance of group discussions

and presentations. It should be noted that within the scope of formative performance assessment component- portfolio- in upper-intermediate and advanced level of the preparatory program, students' speaking skills are tested in the classroom through these task types. Based on teacher response it can be said that despite the fact that the TRACE does not test speaking, this did not lead to a narrowing of curriculum leading to changes upon teachers' methodology. However, 30% of the teachers indicated that materials did not encourage students to make use of the information given across sources in the speaking activities.

Teachers' opinions regarding grammar and vocabulary materials are listed below.

Table 29

Frequency and percentages for teacher perceptions of materials catering for grammar and vocabulary

Materials Covering Grammar and Vocabulary	Never		Sometimes		Frequently		Always	
	f	%	f	%	f	%	f	%
16. Our materials covered grammar & vocabulary that are important for learning English	0	0	3	14	13	62	5	24
17. Our materials covered grammar & vocabulary that are important to do well on tests.	0	0	4	19	9	43	8	38
18. There was a variety of grammar & vocabulary activities	0	0	5	24	14	67	2	10
19. (Grammar & vocabulary lessons were followed by different activities such as:								
19.1 multiple choice	0	0	5	24	10	48	6	29
19.2 matching	0	0	4	19	12	57	5	24
19.3 fill-in the blanks	0	0	2	10	13	62	6	29
19.4 using newly learned grammar and vocabulary in writing	0	0	4	19	11	52	6	29
19.5 using newly learned grammar and vocabulary in speaking	2	10	8	38	8	38	3	14

Large proportion of the teachers (86%) affirmed that course materials frequently covered grammar and vocabulary that were important for learning English. Similarly, it was suggested that the content frequently covered grammar and vocabulary that would help students perform well on the test (81%). 77% of the teachers agreed that there was variety of activities to teach grammar and vocabulary ranging from fill-in-the blanks (91%), matching (81%), using

newly learned structures and vocabulary in writing (81%) and multiple choice (77%). However, using newly learned vocabulary and grammar in speaking generated a variety of response since some teachers suggested this was never covered (%10) and at times covered (38%).

In sum, findings revealed that teachers had positive perceptions overall about the course materials and they believed that materials covered skills and language that would help their students to perform well on the test. However, it was also stated that reading, listening and writing materials were unable to cater for summarising, paraphrasing and synthesizing information from different texts to answer questions. Overall, there was positive washback of the TRACE exerted upon the materials but more than half of the teachers pointed out that the above mentioned skills which could facilitate learning were excluded from the materials.

4.2.4. Findings of the teacher questionnaire for instructional sensitivity

All teachers reported that they think students learned how to find the main and supporting ideas. Also overwhelming majority is of the opinion that students progressed in understanding references (95%), identifying specific information by using reading strategies (90%), and finding out the gist of the text by the help of contextual clues and visuals while reading. Whereas, some teachers disagreed that students learned certain reading skills such as relating ideas to see organisation (29%), recognizing different ideas to understand overall point (24%), reading between the lines (24%), and selecting and making notes from readings (34%) which fostered critical thinking and autonomous learning. In addition, there was a variety of teacher response in terms of whether students had learned integrating ideas across sources into their own writing because 48% of the teachers disagreed and 52% indicated agreement. In other words, nearly half of the respondents affirmed students had not progressed in terms of discourse synthesis. It should be noted here that this finding revealed the discrepancy between the washback intended by the TRACE and actual washback in the classroom.

Table 30

Teacher perceptions of how much students learned in terms of reading skills

	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
13. Find out the main idea(s) of the text.	0	0	0	0	14	67	7	33
14. Identify supporting ideas, details and examples.	0	0	0	0	13	62	8	38
15. Find out the message of the text by looking at visual (layout, headings) and contextual clues (choice of words, grammar structures, examples given) in the text.	0	0	4	19	13	62	4	19
16. Identify specific information by using some strategies (such as reading carefully, taking reading notes, leaving out some information if I didn't understand and tried to guess the meaning, searching for info quickly).	0	0	2	10	16	76	3	14
17. Relate ideas in a reading to one another by understanding how the text is organised, whether ideas follow one another, whether the writer explains his ideas by explanations and examples.	0	0	6	29	14	67	1	5
18. Guess the meaning of unknown words using their knowledge.	0	0	4	19	12	57	5	24
19. Read between the lines to understand the author's attitude, tone and purpose that are not openly stated in the text.	0	0	5	24	11	52	5	24
20. Understand words and phrases such as "This shows …, These are…" that refer to objects, people or concepts.	0	0	1	5	13	62	7	33
21. Recognize different organisations of ideas (definition, cause and effect, comparison…etc.) in the reading text to understand the author's overall point of view.	0	0	5	24	14	67	2	10
22. Select and make notes from reading texts to answer questions.	1	5	6	29	11	52	3	14
23. Combine their reading notes with other sources to write an essay.	1	5	9	43	11	52	0	0
24. Compare and contrast different texts to understand why they are written (purpose), have different styles (tone, author's attitude), for whom the text is written (audience)	0	0	10	48	7	33	4	19

As for listening skills findings revealed a variety of perceptions both in favour and against students' learning and progress.

Table 31

Teacher perceptions of how much students learned in terms of listening skills

.	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
20. Guess the meaning of new words to understand a listening text	1	5	5	24	9	43	6	29
21. Find out the main ideas based on the clues in the listening text such as organisation of the listening text, examples and key words used and questions asked by the speaker(s).	2	10	12	57	6	29	1	5
22. Identify specific information (by paying attention to names, dates, numbers, signal words such as one of the important problems…, firstly,…etc.)	0	0	0	0	11	52	10	48
23. Understand the relationship between ideas	0	0	5	24	14	67	2	10
24. Understand the speaker's use of stress and intonation to signal different sections of a listening text.	1	5	6	29	11	52	3	14
25. Recognize organisation of different sections of a listening text.	1	5	6	29	12	57	2	10
26. Select and take notes from an organised listening text.	1	5	15	71	4	19	1	5
27. Use their listening notes to answer questions.	0	0	0	0	15	71	6	29
28. Combine their listening notes with other sources to write an essay.	1	5	5	24	15	71	0	0
29. Infer who speakers are, what the situation is, speaker's attitude and purpose	0	0	3	15	13	62	4	19
30. Follow the arguments of talks and presentations if they are organised and on a familiar topic.	0	0	6	29	12	57	2	10

All teachers concurred that students had learned identifying specific information by listening for numbers, dates and signal words as well as using listening notes to answer questions. Also a great majority (81%) agreed that students were able to master inference in listening. However nearly 30% reported negative perceptions regarding students' progress in skills such as understanding speaker's use of stress and intonation, recognising organisation, guessing the meaning of unknown words while listening, following arguments of talks, and combining listening notes with other sources to write an essay. In addition, the results demonstrated that a large proportion of teachers did not deem student progress with regard to

selecting and taking notes from an organised listening text (76%) and find out the main ideas

(67%).

Regarding how much learning took place in writing teacher perceptions of certain

writing skills were highly positive and certain skills were regarded as lagging behind.

Table 32

Teachers' perceptions of how much students learned in terms of writing skills

	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
16. Give short answers to the questions about the texts I read or listen to	0	0	1	5	13	62	7	33
17. Complete missing info about the texts I read or listen to	0	0	1	5	15	71	5	24
18. Take organised notes while listening	0	0	3	14	14	67	4	19
19. Take organised notes while reading	1	5	10	48	7	33	2	10
20. Organise their ideas in an essay, generate content, link their ideas to one another in paragraphs and in the whole essay by explaining, using linking words (e.g. In addition, on the other hand, however, furthermore…etc.) and giving examples.	0	0	1	5	16	76	4	19
21. Generate content, organise their ideas in to a plan/an outline, introduce the topic and relate their ideas to each other by using linking words	0	0	2	10	16	76	3	14
22. Support their ideas using explanations, descriptions, statistical data, information from other sources and examples	1	5	4	19	14	67	2	10
23. Conclude their essay by summarising main points and offering their final comments on the topic.	0	0	1	5	14	67	6	29
24. Choose appropriate organisational patterns (such as defining concepts, comparing and contrasting, describing cause and effects …etc.) to develop their ideas in the paragraphs	1	5	4	19	13	62	3	14
25. Write essays of about 400-450 words in which they can organize and express their ideas well and use information from other sources.	0	0	2	10	15	71	4	19
26. Summarise or paraphrase information from other sources such as articles and lectures	1	5	10	48	10	48	0	0
27. Re-write their essay by getting feedback from their class mates	2	10	5	24	12	57	2	10
28. Re-write their essay by getting feedback from their teacher	0	0	1	5	15	71	5	24
29. Edit their own work by self-evaluating it	2	10	4	19	13	62	2	10

	Strongly Disagree		Disagree		Agree		Strongly Agree	
30. Write in an academic style by using mostly correct grammar, spelling and punctuation, as well as variety of grammar structures and appropriate vocabulary.	0	0	2	10	16	76	3	14

Teacher response elucidated that except for using dictionaries effectively and using newly learned vocabulary in their speech, students had learned the grammar and vocabulary objectives of the course. To illustrate, more than 81% reported that students learned how to record and remember newly learned lexis and use these in their writing. Also, 77% pointed out that students could use different grammar structures accurately. However, using newly learned vocabulary in oral production was deemed insufficient by 48%.

Table 33

Teacher perceptions of how much students learned in terms of grammar and vocabulary

	Strongly Disagree		Disagree		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%
6. Record and remember newly learned vocabulary items in an organised way	0	0	3	14	14	67	3	14
7. Use newly learned vocabulary in their speaking	0	0	10	48	7	33	3	15
8. Use newly learned vocabulary in their writing	0	0	2	10	15	71	3	15
9. Use dictionaries effectively	2	10	9	43	8	38	1	5
10. Use different grammar structures accurately	0	0	4	19	14	67	2	10

4.2.5. Highlights from teacher and student questionnaires regarding instructional sensitivity

Comparison of teacher and student questionnaires highlighted several discrepancies especially regarding instructional sensitivity. To begin with 34% of the teachers disagreed that combining reading notes with other sources to write an essay was a learning outcome for the students whereas only 17% of the students indicated that they did not learn this skill. Overall students deemed their performance higher in terms of synthesising information from a variety

of sources in to their written outcome than how participating teachers perceived them. Similarly, as for taking reading notes student response yielded more positive perceptions in comparison to teacher evaluation of their performance in selecting and making reading notes.

There was a dramatic difference between teachers and students regarding how well students learned how to compare and contrast different texts critically to read between the lines and infer the tone, audience, source and purpose. 48% of the teacher response disagreed that students learned to master this reading skill in contrast 18% the students. Contrary to the teacher response, 60% of the students indicated that they learned about texts analysis to identify such inferences.

Similarly, there was diverse teacher and student response regarding how much learning happened in taking organised notes. 71% of the students agreed that they learned how to take organised notes whereas 53% teachers disagreed. It seems that teachers did not evaluate student performance in note-taking as sufficient in contrast majority of the students who deemed that they learned to master this skill.

Student and teacher perceptions displayed variety with regard to how much students learned to use newly learned grammar and vocabulary in their speech. 87% students reported that they learned to make use of newly learned lexis and grammar structures while speaking. However, 48% of the teachers disagreed that students were able to transfer newly learned language into their speech.

Finally, teacher and student response yielded diverse perceptions about how much learning happened in terms of summarising and paraphrasing information across sources. 53% of the teachers stated that they did not think that students learned how to summarise and paraphrase information from different sources. On the other hand, 48% of the teachers expressed positive perceptions indicating that students were able to summarise and paraphrase.

There was diversity between teacher responses. However, 76% of the students indicated that students learned summarising and paraphrasing.

4.2.6 Findings of the interviews with students of PEP for washback and instructional sensitivity

4.2.6.1 General findings

In the first part of the focus group interviews students were asked to focus on instructional sensitivity. In other words, they were asked questions which surveyed their perceptions regarding their progress in the Advanced and upper-intermediate English preparatory courses. When respondents ($n=47$) were asked whether they think they had improved their grammar, vocabulary, reading, listening and writing skills majority stated that there was improvement in terms of language competency and skills.

Table 34

Student perceptions regarding instructional sensitivity of TRACE- Student improvement

	f			%		
	Yes	No	To some extent	Yes	No	To some extent
Reading Skills	30	7	10	64	15	21
Listening Skills	40	0	7	85	0	15
Writing Skills	43	0	4	91	0	7
Grammar	47	0	0	100	0	0
Vocabulary	47	0	0	100	0	0

%64 of the respondents thought that their reading skills improved in upper-intermediate and advanced levels of the preparatory program. However, 21% of the students claimed that when they read a text identifying the main ideas could be a major problem. Furthermore, it was suggested that differentiating between the main and supporting ideas was much more challenging since this required detailed understanding of the text. Additionally, students claimed that were reading to locate the answer to a question rather than critically analysing the ideas presented in the text. 36% of the students stated that they didn't consider themselves

progressing in reading skills because there wasn't an explicit skills training which built upon their fundamental knowledge of 'read the first and the last sentence of a paragraph and you will find the main idea'. 21% of these students stated that they improved somehow because there was a lot of practice and they discovered their own ways of reading more efficiently. When students were asked what they think they learned in terms reading skills in the course and if they can provide some examples 53% of the response indicated that they learned how to find the main idea. 32% of the students mentioned they learned vocabulary and 11% stated that they learned guessing meaning from context and 9% claimed that they learned inferencing, and skimming and scanning. They reported that overall they perceived an improvement their reading skills but had difficulty in defining which reading skills they felt improvement.

In terms of progress in listening skill, 85% of the students considered that they displayed progress. However, 15% of the respondents expressed that there was a lack of explicit training in listening and note-taking in the program but they somehow improved their listening.

Regarding progress in writing, 91% the respondents held the opinion that they improved in terms of writing. Whereas %7 of the students thought that they improved to some extent. 51% of the participants reported that they learned how to organise their essays. 23% of the respondents claimed that they learned using newly learned grammar structures and 17% suggested they learned about using vocabulary in their written production. However, only 10% of the respondents stated that they learned how to correct their errors based on teacher feedback. Furthermore, 17% of the students mentioned that they learned about how to support their main ideas during the course.

Student interview results confirmed the data captured in the student questionnaire regarding instructional sensitivity since majority of the students stated progress in their skills and language after instruction in the PEP. However, based on students' verbal accounts it should be noted that students claimed progress mostly in writing and listening. Some pointed out to a lack of progress in reading saying that the program excluded explicit skills training and students

relied on exam taking skills (e.g. students were reading to locate the answer to a question rather than critically analysing the ideas presented) to cope with reading skills rather than deep learning.

Washback on materials

In an attempt to examine students' perceptions as to washback of TRACE, in the next part of the interviews students were asked questions which aimed at eliciting their ideas regarding the washback of TRACE on teaching materials and the methodology employed by their teachers. To unveil student attitudes to teaching materials used in class, respondents were asked whether they think course materials were effective in terms of their learning. Views of the respondents are summarised in the table below.

Table 35

Student views regarding contribution of course materials to learning English

Course Book (f)			Supplementary Materials (f)			Vocabulary and Grammar Booklet (f)			BLC (f)		
Yes	No	To Some Extent	Yes	No	To Some Extent	Yes	No	To Some Extent	Yes	No	To Some Extent
12	26	9	39	2	6	43	4	0	12	29	8

26% of the students regarded that the course book contributed to their learning because topics were interesting, and reading and listening texts presented in the book were perceived to be useful. However, 55% of the respondents stated negative perceptions about the course books used in both levels since the follow up exercises were not compatible with the exams. It was claimed that most of the exercises focused on open-ended questions which were not asked in the exam leading to a very frequently voiced concern; "It was not preparing us for the exam". Also another frequent comment was that course book was often supplemented with extra

worksheets given by the teachers and therefore students didn't feel that they were doing the course book. Many students commented that they had not written anything on the book and not revisited it for further practice.

With respect to supplementary materials 83% of the respondents expressed positive perceptions. It was mentioned that supplementary materials were deemed as the most useful part of the course because it resembled exam and they provided students with the opportunity of exam preparation.

As for the Blended Learning Component (BLC) of the advanced and upper-intermediate courses, 26% of the respondents claimed that these materials contributed to their learning English since BLC tasks were in the form of quizzes with exam-type multiple choice questions and they recycled grammar and vocabulary. Some students stressed that BLC had contributed a lot to learning vocabulary because target vocabulary was recycled through these exercises. However, for 62% of the students BLC was inefficient because:

(1) Materials were not user friendly since they needed to scroll up to read the text and then scroll down to look at the question.

(2) Students were denied of instant feedback to their response to open-ended questions included on the materials.

(3) These materials were not revisited in the class.

(4) Materials were difficult.

(5) Materials were not interactive and therefore students perceived them as worksheets given online. Thus, they wanted to do them in hard copy instead of working on soft copy.

(6) There were some technical hitches in downloading the texts, uploading answers and listening to recorded materials.

Additionally, the respondents were asked to comment on the difficulty level of the course materials and were requested to express their ideas about the course book, supplementary materials, vocabulary and grammar booklets and BLC.

Table 36

Students' views regarding difficulty level of course materials

Course Book (f)		Supplementary Materials (f)		Vocabulary and Grammar Booklet (f)		BLC (f)	
Yes	No	Yes	No	Yes	No	Yes	No
27	20	11	36	5	42	30	17

When students were asked whether they think that the course materials prepare them for the TRACE their response was mostly positive for supplementary materials and vocabulary and grammar booklets. However, many respondents stated that that they did not think that course books prepared them for TRACE exam. There were variety of responses regarding BLC among the students indicating that the majority did not perceive BLC efficient in providing preparation for the exam.

Table 37

Student perceptions as to whether course materials prepared them for the test.

Course Book (f)		Supplementary Materials (f)		Vocabulary and Grammar Booklet (f)		BLC (f)	
Yes	No	Yes	No	Yes	No	Yes	No
13	34	47	0	47	0	15	32

Consequently, it can be concluded that the TRACE and its requirements had an immense effect on students' perceptions of course materials. Students deemed in-house prepared supplementary materials as the most useful to their learning because they were closely aligned

with the requirements of the exam regarding content, text types, task types and format. TRACE had a positive washback effect on course materials.

When respondents were asked to think in retrospect about reading, listening and writing activities and tasks that were done in their classes all students recited essay writing, reading, lecture listening and responding to follow up questions. However, some respondents mentioned an array of activities involving jigsaw reading, discussing answers to reading texts in groups (11%), linking readings to writing (11%), vocabulary games and exercises (26%), speaking activities (11%), summarising (13%), role-play (4%), activities that involved integration of technology (13%). However, students claimed that frequently implemented tasks in the classroom tended to point to practice in exam type tasks rather than an even balance of four skills. In this respect student response in interviews was contradictory to student response captured through the questionnaire since the former implies narrowing of curriculum down to tested skills because instruction in the classroom excluded speaking activities and other communicative and collaborative learner-centered tasks.

Next, students were asked which of these tasks were believed to be directly related to the test and may help improve their scores. Their response revealed that exam practice, example TRACE, essay writing, reading, listening, vocabulary work and all tasks done in class were seen as tasks which are directly related to the exam and would help students contribute to gains in scores. In addition, 96% of the students claimed that course content and content of the exam were similar because exam questions and tasks types were similar. All respondents declared their belief that course content will help them be successful in TRACE. Some also commented that course content would help them in gains in scores not only in TRACE exams but also in all other prior and prerequisite exams (formative and summative) they need to take in their levels before TRACE.

Finally, students were asked about their ideas about how well they think the course supported their students to learn English and be successful on TRACE. All respondents stated preparatory course supported them a lot to be successful on TRACE. However, indicating the negative washback of the TRACE, some students expressed that it did not help them improve their level of English.

4.2.6.2 Emerging themes

Data analysis of interview transcripts revealed three categories which involved exam-orientedness, materials induced factors, and teacher induced factors. Table 38 represents the frequency of these categories.

Table 38

Frequency of emerging themes and sub-themes of student interviews

	f	%
Exam Orientedness	31	74
Materials induced factors		
Discourse synthesis and integration of skills	19	40
Consistency between technology integrated learning materials and the exam	10	21
Topic familiarity	14	30
Teacher Induced factors		
Effect of teacher's practice- Variation between teachers	22	47
Coaching	16	55

Exam Orientedness

Data analysis indicated that the TRACE had a negative washback on student learning since students were inclined to favour activities intended for test orientation and coaching. One of the factors behind excessive exam-orientation of the learners was reported to be their previous educational background. Most of the respondents concurred that they got used to multiple choice high stakes testing culture which displays significant differences in comparison

127

to academic demands of the PEP and the university. However, as a high-stakes proficiency test, TRACE was similar to their previous educational experience which relied on the dominant multiple choice testing procedure. Consequently, their response revealed that majority of the students take requirements of the exam to the center in their learning process and approach and focus on exam-taking strategies rather than learning.

> *We are used to copy and paste culture. I mean we assume that an answer to a question is there in the text, sitting still. As if we need to quote a sentence from the text and copy it as the answer. But actually, what we are asked to do is not copying and pasting. We are requested to evaluate the idea critically and then give the main idea.*

Along the same line, in response to the interview question which surveyed whether students think that they improved their listening skills in the course, one of the respondents highlighted the exclusion of learning English during high school saying:

> *Frankly, I think that I have improved myself. Previously at high school and in our schools we hadn't studied English like we do now. I mean we hear English in all our lessons. At times we watch videos and watch movies. I really think that these contribute a lot to our learning.*

It can be inferred that previous educational background reinforced rote learning and had not fostered communicative and creative language learning through authentic materials.

When students were asked to self-evaluate themselves and share their perceptions regarding whether they improved in the course with respect to language and skills, there were comments focusing on strategies to get better grades on the test rather than critical self-evaluation. It was concluded that the students were highly exam-oriented and lacked awareness on how they progressed in the PEP. Instead of approaching efficient learning as a whole, some respondents degraded the concept of learning a foreign language to learning particular skills. To illustrate, one of the respondents highlighted the intimate relation between progress and improvement of test scores by drawing attention to association between how progress in vocabulary knowledge would benefit test scores.

My vocabulary knowledge was very weak before I came to prep program. Actually I think that I have improved because we learned vocabulary that would come up in the reading parts of the TRACE. As a result, we can do the readings much more easily. We also use these vocabulary items in writing and this brings about higher grades.

Therefore, it can be stated that majority of the students were more focused on test taking skills and strategies for increasing their gains in scores rather than a deliberate focus on learning. Similarly, another response underlined the assigned value to memorising sentences which would display uses of level-specific grammar objectives ('complex grammar structures') in essays in getting high scores from the exam: *"If we learn one or two new structures and use these in our writing we would not have any problems in grammar. Listening has no grammar and reading is based on your vocabulary knowledge. So it is enough to make use of couple of complex grammar structures in writing".* It is significant to note here that according to some students, progress in the course was evaluated by the scores achieved in the tests and the focus was shifted towards test-wiseness rather than focus on progress of language and skills. The formula of getting high scores in exam was prescribed as memorising newly learned vocabulary items as well as grammar structures and putting some of these into use in the writing part of the exam.

In a similar line of thought when asked to comment on their progress in listening, instead of commenting on skills that they learned, some students chose to talk about the exam skills gained and included ways of getting higher scores in listening in their response:

Note-taking was bad for me. I learned what could be asked in exam because there are certain things that we need to take notes of. I directly take notes. The speaker at times emphasizes important information. For example, if there is a number s/he says that by stressing that information. So, I directly write that part. There can be information that we miss but as long as I understand the overall lecture I can do some questions even by using logic.

Findings of the interview data of the study pointed out to a negative washback effect on student learning since students were inclined to be test-oriented and perceive activities oriented

towards the test or test-specific coaching to prepare for the high-stakes proficiency exam. This finding was in line with the contention of other washback studies that students attach importance to skills tested and focus highly on exam-related activities, test content and format (Gosa, 2004; Stoneman, 2006; Shih, 2007, Tsagari, 2009).

Materials related factors

Majority of the students (72%) stated that they were holding negative perceptions towards the course books that were used in upper and advanced levels because the materials were not seen compatible with the exam. In other words, the respondents stressed that there was a mismatch between the content of the instructional materials and that of TRACE with regard to question types that followed the oral and written texts, lengths of the oral texts and number of the reading texts. Since the TRACE had four reading texts in the reading section, students expected to see multiple reading texts in their course materials. Some of the comments highlighted students' tendency to evaluate learning materials as efficient and conducive to learning because they included of multiple choice test tasks: *"There are no tests in the course book. We generally work towards the test but the book doesn't have any tests"*. In addition, there were comments regarding the difficulty level of the reading questions that were included in some follow up tasks in the book.

> *I didn't find the book effective. The topics were fine. The book focused on a variety of topics but the questions need to have more details. When I compare the questions in the book with the questions on the exam I see that answers are quite obvious. Questions in the book are a lot easier. What does the writer want to say in here? I mean we could eliminate the two sentences (means distractors-) so we were left with the final and last option. In this respect I didn't find the book sufficient.*

While some students regarded the book inefficient by totally disregarding whether it helped them to learn English, some others differentiated between learning English and getting ready for the exam. Some response stated that the course book was useful with respect to learning English but inefficient in terms of getting ready for the exam. All students stated that supplementary materials were beneficial for their learning English as well as getting prepared

for the exam because they apparently found consistency between the supplementary materials and the exam: *"The papers…worksheets you have given to us were beneficial because they are geared towards the exam. They even have the same form for instance the question types…even the font of the materials is the same".*

Consistency between technology-integrated learning and the exam

Data analysis revealed that for 21% of the students TRACE exerted some effect on technology integrated course materials which were utilised within the scope of BLC at the advanced and upper-level in the PEP. These course materials were uploaded on course management system of the PEP for student use. Student response indicated both negative and positive washback of TRACE on BLC materials.

It was suggested that task types and question types of some of the BLC materials and those of the test were identical. In other words, it was indicated that some BLC materials lacked a variety and resembled the test with respect to its design and question types.

> *BLC questions are multiple choice, just like in the exam. But my reading is worse when I read texts on the internet because I need to underline. I think that it would be better, if this reading homework is given as hardcopy. Then I can take reading notes. I can underline the main ideas. But I cannot do these on the internet. As a result, I get worse results. I have more wrong answer. But if I do reading on paper my scores are better.*

It was observed that students treated BLC tasks as exam practice on-line and welcomed BLC tasks that echoed the requirements of the exam. Consequently, the students did not appreciate BLC materials when they were asked open-ended questions eliciting a range of student response rather than just multiple choice questions. Some students expressed their disillusionment with BLC materials. One respondent stated: "The questions in BLC and the reading questions in the exam are very different".

However, positive washback was also mentioned since some BLC materials were regarded to be closely related to other course materials, reinforcing course objectives. The following is what one of the respondents stated *"I think BLC and the exam are related to each*

other because vocabulary and listening reading exercises in BLC and exams are the same. I mean they complement each other. They built on each other."

Topic familiarity

While evaluating course materials and their contribution to students' learning, many students mentioned on the effect of topic familiarity in these materials. Some materials were deemed above their level because they included unfamiliar topics. However, students regarded these materials beneficial saying that these topics may appear in TRACE, *a*s manifested in the following response: *"I think our course book is difficult but there are units on urban planning, innovations, and art. I don't know much about these but these topics can be asked in TRACE so I think they are beneficial".* Many respondents stated that topic familiarity was important for their performance in exam and they believed that course materials prepared them for potential exam topics to appear in TRACE as reflected in the following student response.

> In TRACE, all exam was on one topic; environment. I don't know; there are advantages and disadvantages of this. For example, you see the vocabulary items that you can use in your writing in reading texts. You can keep that vocabulary in mind. But if it is an unfamiliar topic it is very difficult. If you ask a girl to write about football she will have difficulty. Your English level may be good but topic is difficult. Then, how well can the exam assess your performance? I am not sure. When we read about a topic in our course book and materials and discuss then dealing with that topic becomes easier.

Discourse synthesis and integration of skills

40% of the respondents indicated that TRACE impacted materials-especially in-house prepared supplementary materials because task types, question types, and genres and length of the texts resembled those in the exam that facilitated integration of skills as well as discourse synthesis.

> *I think worksheets are very good. In each material our focus changes based our needs. If we need more reading or listening practice or grammar our teachers caters for that. In those materials, there are background information which supports our understanding, stuff from internet. Also our teacher may start a discussion based on information from internet.*

Findings also may imply that discourse synthesis had a facilitative effect on their learning. *"I think we had written many writings but if we wrote 10 essays I was able to use information from different sources towards the end of the module when I learned grammar and vocabulary better".*

Some students indicated that discourse synthesis and all classroom practice that aimed at further practising this skill raised students' awareness and helped them develop the competency of using information across different sources into their written outcome.

> *Also now we're doing this involuntarily. From the top of my head, let's say the topic is art. We first do the reading and then listening exercises. Then, we write essays based on those. But when I first entered TRACE, I didn't pay attention to the sentences in the reading. I wasn't reading it carefully. I was just reading it for the questions of the reading. But now, if I am given a paper, I will know that it will be followed with other activities. I will read the sentences carefully, I will do, you know. Because it will affect how I write my essay too.*

Many respondents referred to the positive washback of the thematically integrated TRACE because it raised students' awareness to generating, organising and linking ideas as well as modelling vocabulary and sentence structures. It was frequently mentioned that supplementary materials, which were also designed similar to those in TRACE specifications encouraged students to synthesize ideas across a variety of sources and had a positive effect on students' learning.

> *Especially in writing... You know those bunch of papers (supplementary materials), I love those. During writing this essay my best friends were those bunch of papers. I mean, when the instructor says skim and scan the main ideas, get better in writing accordingly, later you will write one without having a look at those papers. I studied accordingly, I started writing accordingly. I even copied some sentences from those. After that I started to get better and better. First, learn what it is, then, you have the inspiration and creativity kicking in.*

Discourse synthesis has been deemed as a facilitator for students' language growth in teacher interviews. However, when teachers and students were surveyed through

questionnaires about how much learning has happened with regard to synthesising ideas across different sources and integrating these into one's own work, both parties concurred that students had difficulty in mastery of this skill. It was also noted that classroom materials excluded tasks reinforcing synthesis of ideas. In other words, data analysis revealed that although participants perceived discourse synthesis as 'beneficial' to learning, they also reported that students could not acquire this skill satisfactorily.

Teacher Induced factors

Data analysis of student interviews indicated that the following e teacher induced factors were related to washback effect of TRACE.

Effect of teacher's practice- Variation between teachers

47% of the student response indicated that teachers' classroom activities displayed variety regarding the use of different media and modality. It was reported that in some classes students were exposed to a range of authentic materials including short videos, talks and texts from websites in an integrated manner, where as in some others there was no sign of an integrated approach or exploitation of different media and modalities. In other words, it was stated that only some teachers provided students with a number of oral and written texts in order to aid them in learning about the topic (idea generation) and practising their language and skills. It can be inferred that teachers' understanding of the exam constructs and underlying principles, including integrated nature of the exam and discourse synthesis, had a facilitating effect on instruction and this assessment literacy reflected in teachers' methodology.

> *We don't only do listening exercises. We also watch videos to understand the topic better. This also helps our organisation in writing because we get different ideas. Our teacher gives us links to websites for practising our listening. These were extra work. I sometimes listened to these at home and they helped me a lot. I think this is my own effort as well as my teacher pushing me.*

Individual teacher's knowledge about and perceptions of exam regarding tested skills, question formats, and exam specifications affected their instruction. Respondents provided comments similar to the following:

> *Our teacher was a new comer to the school and she didn't know much about the system. So it was a bit chaotic for us. We did the course book and we were not given any photocopies. When I talked to other friends in other classes I learned that there were many worksheets, note-taking sheets. I first saw the note-taking and listening to a lecture in Mid-Module achievement test and we were shocked.*

Additionally, some students mentioned that treating the instructional materials was diverse. In some classes it was a common practice to do the readings and listening under exam conditions regarding the seating and timing, while in some others, there were more communicative activities with pre-teaching vocabulary, lead-in, and tasks to build background information, guidance in skills training to deal with the materials. Some students indicated that they learned better when they discussed why their response to a multiple choice question was correct or wrong. In addition, contrary to exam practice which diminished student-to-teacher and student-to-student interaction in class, some teachers were reported to employ more communicative tasks and to provide explicit strategy training. One respondent stated:

> *I used to write words in isolation when I took listening notes. Then my teacher*
>
> *shared her own notes and made us share ideas about how to take notes which would*
>
> *give me the context and made me remember the content of the lecture. I also learned*
>
> *to use arrows between ideas to see the connection*

Based on students' responses, it can be inferred that for some teachers TRACE exerted a negative washback. The students suggested that TRACE changed not only what teachers taught but also how they taught. While it was also suggested that TRACE affected some other teachers at varying degrees since even though they were prepping their students for a high-stakes proficiency exam they did not change how they teach.

I can say that I learned to read better by making many mistakes in my responses to the reading questions. In the class our teacher was always warning us saying that the required answer and we were giving general answers. I learned better when our teacher made us discuss our answers and showed us what was the answer and why it was the right answer.

This comment was not in line with the findings obtained from student and teacher questionnaires where it was mentioned that the washback effect of the TRACE may often lead to a change only in content (what teachers teach) and may not exert effect on methodology (how they teach).

Coaching

For more than half of the students (55%), especially for those at the advanced level, the focus of the instruction in the classroom depended on their teachers coaching them for the TRACE exam. Students distinguished between Advanced Level and upper-intermediate level by indicating former as a course which prepared them for the TRACE exam and the latter as a course which focused on communicative English*: "In upper-intermediate we added more information upon our existing knowledge of English. I think that Advanced is a course which teaches how to get good grades in the exam. We are learning strategies mostly. Upper is more related to increasing grammar knowledge and learning vocabulary".* One of the students stated: *"I think that teachers teach in a tactical way by explaining you should pay attention to this and that"* and put forth his belief that these "tactics" would raise their exam performance, leading to better test grades.

In addition, it was mentioned that some teachers exploited materials with an explicit focus on how to increase gains in test scores by making use of content and language from different sources saying: *"At times our teacher found some informative videos and we listened to these videos. Then our teacher would write down the important vocabulary items saying that you can use these words and specific topic related collocations in your writings".*

The best side of my teacher was that when I wrote a sentence my teacher would give me 2-3 example sentences saying that you can express your idea in better words and grammar in this way. She did this to all students. Then we were able to build better and more elaborate sentences in the writing part of the exams. This increased our grades in return and I am sure this will help us in TRACE as well.

Consequently, it can be inferred that the TRACE exerted negative washback on teaching because some teachers were coaching the students to increase their gains in scores instead of getting involved in meaningful learning.

4.2.7. Findings from interviews with language instructors of PEP for washback and instructional sensitivity

4.2.7.1. General findings

Interview questions mainly surveyed teachers' perceptions regarding instructional sensitivity and washback of TRACE in relation to teaching materials and classroom activities. In the former area, questions were asked in an attempt to examine how much learning took place in terms of reading, listening and writing. In the latter section, teachers' attitudes towards teaching materials (including course book, supplementary materials, vocabulary and grammar booklet, Blended Learning Web Materials) and tasks and correspondence between teaching-learning and being successful on TRACE were focused on.

Findings related to instructional sensitivity

In the first part of the one-on-one interviews, the focus was on instructional sensitivity. In other words, teachers were asked questions which surveyed their perceptions regarding students' progress in the advanced and upper-intermediate English preparatory courses. When respondents ($n=14$) were asked whether they think their students had improved their ability in grammar, vocabulary, reading, listening and writing skills, majority stated that there was improvement in terms of language competency and skills. Table 39 below outlines teacher perceptions of student improvement in skills and language during these courses.

Table 39

Teacher perceptions regarding instructional sensitivity of TRACE- Student improvement

	f		
	Yes	No	To some extent
Reading Skills	10	2	2
Listening Skills	10	4	0
Writing Skills	11	1	2
Grammar	14	0	0
Vocabulary	14	0	0

Most of the respondents (%71) think students' reading skills improved in upper-intermediate and advanced levels of the PEP. However, some teachers (29%) commented that students can get the answers to reading texts accurately but this doesn't mean that they actually comprehended the main ideas. Some teachers attributed this to students' inclination to be test-oriented and read to locate answers instead of reading critically for the main ideas. Whereas, for some, this could be associated with weak items in the test that could be answered without comprehension. It was observed that students read the text and use their grammar and vocabulary knowledge alongside their test-wiseness in order to find the right answers to the questions. Consequently, it was reported that students have partial understanding and ineffective reading comprehension. Furthermore, most respondents agreed that students have difficulty in summarising as well as finding answers to the questions which ask 'how and why'. According to some teachers, students find inferencing and reading for main ideas challenging because of their lack of motivation in reading in L2 as well as L1 and lack of extensive reading outside class.

In terms of progress in listening skill, majority of the teachers (71%) considered that their learners displayed progress in their listening skills. However, 29% of the respondents expressed that student learning was impeded due to (1) insufficient amount of listening materials, (2) discrepancy between in class practice listening tasks and listening tasks used in exams regarding difficulty level, (3) lack of while listening materials in the course, (4) lack of strategy training in listening, (5) low motivation of the students, (6) length of lectures, (7) speed

of lectures and (8) low level of student comprehension due to low level of grammar competency and inadequate knowledge of vocabulary. It was also suggested that at times students were able to respond to listening questions that follow a lecture without taking notes by using their test-taking skills and strategies.

Regarding students' progress in writing, 79% of the teachers held the opinion that their students improved their writing ability whereas %14 think that they improved to some extend and %7 think that they haven't improved at all. Most of the participants who reported negative perceptions indicated that students had difficulty in generating ideas and providing logical connections that weaved main ideas together.

> *What I observed in students' writing, the most common problem I observed in addition to using a very limited grammar and very limited vocabulary was something related to their cognition. Their way of presenting the ideas, supporting them and explaining the ideas. It was actually very weak in terms of logical connections. That's what they lacked. They thought that they were actually supporting ideas, whereas, when you question their reasoning in your feedback, they kind of also couldn't see themselves how the supporting sentence was actually related to their main idea. So, that was the main problem that I saw. Being able to see very clear and strong links between ideas that are really relevant to each other was rare. There was quite amount of irrelevance of all kinds and very farfetched examples. What idea really is supported by what kind of detail? What kind of support is really a logical connection between ideas? What is a cause-effect relationship like? What is a problem solution relationship like? That was what we needed to focus on, because it was there sometime, even it was there, they were not aware of it. So, it was evident in lack of use of appropriate linkers, actually.*

Furthermore, it was often mentioned that students didn't benefit from teachers' written feedback because they felt discouraged by use of error codes, amount of correction and they resisted reading feedback.

In an attempt to examine teachers' perceptions as to how much learning happened during the courses, respondents were surveyed about their perceptions regarding their students' progress in reading, listening and writing skills in the course. Their response is outlined in the table below.

Table 40

Teacher perceptions regarding instructional sensitivity of TRACE- Skills

Reading	%	Reading	%
Finding specific details	50	Critical thinking	14
Inference	50	Analysing the text	7
Guessing vocab from context	36	Finding main idea	7
Search Reading	29	Differentiate between main and	7
Awareness of reading skills	29	supporting ideas	7
Cross textual analysis	29	Improve reading speed	7
Improve reading strategies	29	Understand texts	7
How to deal with long texts	29	Commenting on texts	7
Reference	22	Text cohesion	7
Seeing organisation of a text	14	Extensive reading	
Expansion of vocabulary knowledge	14		

Writing	%	Listening	%
Organisation	64	Understanding main ideas	71
Sentence structure-accuracy	36	Note-taking skills	64
Using newly learned vocabulary	22	Listening for specific details	29
Discourse synthesis	22	While-listening	29
Using newly learned grammar	22	Making inferences	22
Supporting main ideas	14	Use of linkers and transition signals	14
Generating ideas	14	Exposure to different accents	7
Justifying ideas	7	Focusing on long texts	7
Error correction based on feedback	7	Organisation of lectures	7
		Guessing vocabulary	7
		Guessing content of the lecture	7

Teachers claimed that the most frequently encountered difficulties in learning seemed to be related to reading for the main ideas, inferencing, summarising, paraphrasing, recognising the link between ideas and low level of grammar competency impeding comprehension. A respondent commented that one of the main problems which students face in terms of reading skills is their preventage from understanding the text because of a lack of vocabulary which impedes 'accessing reading as a whole'.

I think that there has been an improvement in terms of variability to access information more quickly and get the answers more quickly thanks to skimming and scanning exercises done in class. However, I wouldn't say that their general understanding or their global understanding of the text has necessarily improved. They are better getting answers but still not so good at developing a general understanding of the text. I think that if I were to ask them to summarize it, still they would have some problems. I mean they get the main ideas but there are certain points in terms of the details where they don't necessarily understand. They can get

the answers for it but it doesn't mean that they actually comprehend it. Often that's, they will read the text and they use their grammar and vocab knowledge in order to find the right answer to the question but that doesn't necessarily mean that they understand what it means.

On the other hand, with regard to students' improvement of listening skills, the most commonly encountered difficulties were reported to be guessing the meaning of the words, inferencing, listening for specific information and recognising organisation of a text. Also, nearly all teachers stated that there was lack of while listening materials, and therefore, their instruction lacked practice in listening to a text and responding to comprehension questions both upper-intermediate and advanced courses.

Findings related to washback on materials

In an attempt to unveil attitudes to teaching materials used in class, respondents were asked whether they think that the course materials contributed to their students' learning English. Views of the participating teachers are summarised in Table 41 below.

Table 41

Respondents' views regarding contribution of course materials to learning English.

Course Book (%)		Supplementary Materials (%)		Vocabulary-and Grammar Booklet (%)			BLC (%)	
Yes	No	Yes	No	Yes	No	To Some Extend	Yes	No
21	79	100	0	86	7	7	50	50

21 % of the teachers claimed that the course book contributed to students' learning because its content matched curriculum objectives of the courses, topics were interesting, and skills presented in the book were perceived to be useful. However, an overwhelming majority of the respondents (79%) showed negative perceptions about the course books used at both levels, concurring the course books were not effective enough to improve students' English.

A wide array of reasons was outlined ranging from inclusion of unfamiliar, uninteresting, conceptually difficult topics, difficult texts, and short listening texts to lack of variety in reading texts and explicit skills training. However, the most commonly stated reason was that commercial course books were not closely aligned with the content of the test. Many teachers expressed their dissatisfaction for not exploiting the tasks such as responding to open-ended questions suggested in the books because they were not perceived as replicating exam-type tasks and questions by the students. *"We don't use most of the book and the reading and listening texts that are most commonly used from the books are supplemented it with multiple choice questions"*. It was claimed that the mismatch between the tasks, exercises and questions types in course books and those of the TRACE was increasing teacher load by requiring them to develop materials that replicate exam type tasks and questions. The following quote is an example:

> *The main challenge in upper course is that our book doesn't provide enough listening for them and because the listenings, the lectures in the book are shorter than the listening the lectures in MAT and and LAT (formative and summative test of upper-intermediate level). In MAT (Mid-Module Achievement Test) we have at least fourteen minutes of lecture. But in our book, the first lecture is for example only five minutes. So when they sit the MAT exam, they are like shocked because they cannot concentrate enough. After ten minutes they lose everything. And the weakest point in my class in MAT was listening. Out of how many, fifteen questions they were able to do three, four, at most five correct answers. So, I don't think enough learning happens in terms of listening because of our course-book.*

In addition, many teachers didn't consider the course books relevant to students' needs. Nor did they consider the course books in line with course objectives due to lack of focus on grammar teaching, lack of exam type questions, 'irrelevant writing' and 'speaking parts'. Here it is important to mention that respondents regarded writing and speaking exercises of the book as 'irrelevant' in comparison to content of the exam which indicated negative washback in the form of a narrow focus on curriculum.

On the other hand, with respect to supplementary materials, all respondents expressed positive attitude. One of the teachers claimed that the reason for their positive attitude was that supplementary materials had *"face value"*.

> *It looks like something which they would come across in the exam. It also has the same, similar length as listening or reading as they come across in the exam. Therefore, it provides a high motivational value to the students and I find that they're more receptive to supplementary materials. Whereas in our book the texts are generally short, listening are generally shorter and they are of a high level.*

Teachers often stated that supplementary materials that are prepared for the levels took students' exam-oriented expectations, their needs and course objectives into consideration and were tailor made. It was also suggested that supplementary materials echoed the specifications of the exam so well that "supplementary materials give idea about TRACE if you are a new teacher". One of the respondents stated: *"If we didn't have supplementary materials I would be lost"*

> *Supplementary materials were better I guess; because as you know we have objectives in our curriculum and we focus on certain objectives and we assess these objectives in our exams. I think those supplementary materials were especially, I'm not sure about the learning English, but they were I mean helpful for our students to be successful in the exam, mainly; because they focused on the same objectives and they had exam type questions.*

In-house prepared grammar and vocabulary booklets were also perceived positively by 86 % of the teachers. They expressed that booklets provided students with "concrete evidence of what they learned" when they see a list of target vocabulary of the course and the list of grammar topics that would be covered. Additionally, both teachers and students thought these booklets were beneficial because they were deemed relevant to testing as a means of achieving higher scores on exams in the program. After all booklets were seen as a list which guided students into the content of exams. Furthermore, one of the teachers highlighted that vocabulary booklet was useful because it provided skills training in certain aspects of vocabulary learning by giving information about word formation and guessing meaning by examining prefixes and suffixes. However, there were some criticisms especially with the grammar booklet on grounds

of involvement of meta language and difficulty level of tasks which guide students to analyse the grammar structures in use.

As for the BLC component of the advanced and upper-intermediate levels, half of the respondents remarked that these materials contributed to students' learning English since exercises in the form of quizzes and question types reflected multiple choice exam type questions and they recycled course content, grammar and vocabulary. Some teachers stated that BLC was beneficial because it encouraged out of class work, and contributed positively to students' self-studying skills. However, it was frequently noted that these benefits were only for those students who put in effort and did the exercises because they had intrinsic motivation for learning and progressing. In contrast, the other half regarded BLC as inefficient saying that (1) difficulty level of listening tasks and reading texts were beyond students' level, (2) students wanted to do them in hard copy instead of working on soft copy, (3) there was no follow up due to time constraints in the course to exploit BLC, (4) students didn't see the rationale behind BLC and perceived them as a way of gaining scores on classroom participation, (5) there were some technical difficulties with opening the texts uploaded and answering the questions, and (6) students resist doing homework.

Furthermore, the respondents were required to comment on the difficulty level of the course materials for the level. They were asked to express their ideas on the course book, supplementary materials prepared by other teachers, in-house prepared vocabulary and grammar booklets and BLC.

Table 42

Respondents' views regarding difficulty level of course materials

Course Book (%)		Supplementary Materials (%)		Vocabulary and Grammar Booklet (%)		BLC (%)	
Yes	No	Yes	No	Yes	To Some Extend	Yes	No
36	64	71	29	93	7	64	36

64% of teachers considered the difficulty level of the course book above the students' ability level and concurred that difficulty level was leading to not only to comprehension problems but also motivation problems for the students. On the other hand, 36 % of the teachers expressed that they thought that the course book was appropriate in terms of difficulty. In contrast, most of the teachers (71%) stated that supplementary materials prepared by teachers were at the right level for the students. Similarly, in-house prepared materials were regarded to be at the right difficulty level for the students' profile by almost all teachers (93%).

When teachers were asked "Does course materials prepare your students for TRACE?" their response was mostly positive for in-house prepared supplementary materials and vocabulary and grammar booklets. However, many teachers expressed that they weren't of the idea that course books prepared their students for TRACE exam. There were also differing ideas about BLC which are summarised in Table 43 below.

Table 43

Teacher perceptions of Blended Learning Component of the courses

Course Book (%)		Supplementary Materials (%)		Vocabulary Grammar Booklet (%)		BLC (f)	
Yes	No	Yes	No	Yes	No	Yes	No
29	71	100	0	100	0	57	43

Data analysis revealed a positive washback effect of TRACE on supplementary materials and vocabulary and grammar booklet since all teachers claimed that these materials were beneficial for learning. As a result, it could be concluded that the TRACE led to a positive test effect on materials which were designed in close alignment to TRACE specifications and underlying principles except for the BLC materials. Data analysis also revealed that BLC

materials which entailed technology integration have led to both negative and positive perceptions.

Findings related to washback on teaching

Findings revealed that teachers employed a variety of class tasks and activities. Teachers indicated that their repertoire of classroom teaching involved a wide variety of activities ranging from implementing technology-integrated tasks (i.e. using Padlet; online posters to recycle vocab, summarising content of units, and synthesise information across texts), reading individually and checking the answers in pairs and in groups, preparing presentations, writing summaries, practicing essay writing, using newly learned words in sentences, writing paragraphs responding to lectures and using newly learned vocabulary, back-outlining a model essay, asking questions from students to personalise the topic, peer assessing essays, self - assessing essays, discussing responses to lectures, having group discussions, working on grammar exercises, listening to web-based lectures, note-taking, doing while listening activities, asking students to respond to vocabulary quizzes and games , to assigning research projects about the topic of the unit and exam practice among many others.

These reported activities done in class were in line with the findings of the teacher questionnaire and displayed a great variety regarding methodology but few teachers mentioned focusing on integrated skills and employing activities that fostered discourse synthesis.

> *I provided them with one question and they wrote a paragraph about it. For example, in conformity unit, I simply asked, are you a conformist person or non-conformist person? But I do not want any essay here. Just free writing… Just give me one paragraph about yourself. So they can use the related vocabulary in their writing. So, in a way I mixed writing, listening and reading. So I used writing as a follow-up activity after listening and reading.*

Further, a majority of the teachers believed that in class activities would improve students' scores because teaching was geared towards the exam.

> *I think the writing sessions can be given as an example. Because in the writing part you know, they write an essay related to the topic they read in reading and listening.*

So, in our lessons they write an essay related to the topics that we covered in the lesson. Again in listening they take notes and then answered the questions, lectures. What else? Not all the time but sometimes, the reading parts; towards the end of the module, I give the supplementary readings to students as a test. I tell them to imagine they are in the exam, just test yourself don't talk to anyone, don't use your mobile phones and dictionaries to check the meaning of the words, maybe this could be counted as well.

In addition, when teachers were asked whether their selection of activities and tasks were affected by TRACE, they provided positive response. Majority of the teachers (64%) stated that TRACE affected their selection of tasks and 29% of the teachers claimed that it affected them to some extent. Only one respondent (%7) stated that TRACE had no effect on selection of tasks for the class. Most teachers (%64) responded that all tasks done in class were believed to help students perform better in TRACE. All teachers who were teaching Advanced level concurred that exam practice in the form of previous TRACE exams and timed reading, listening and writings were believed to result in boosting up the scores on TRACE. For some teachers, strategy training on how to answer specific exam questions such as cross textual reading questions, and for some teachers using the published exam preparation booklet of another tertiary level institution were assumed to be beneficial.

Definitely, I mean because as I said we need to, they need to pass and complete prep successfully. In order to do that they have, we have an exam reality; proficiency exam, we need to consider that I was of course considering the skills they were going to use while just answering the questions in TRACE, so, yeah I, that was something that I kept in mind.

In response to a set of questions which aimed at surveying correspondence between teaching, learning and success on TRACE, teachers concurred that they think content of the course (what they teach in class) and TRACE are similar. They were similar because both the exam and the course content were theme-based, involved discourse synthesis, catered for the very same objectives with the exception of listening which was excluded from the course content. It was pointed out that the question types used in course materials and those in the exam were the same. Also some teachers commented that topics used in the courses were

chosen for the exam. However, when teachers were asked what changes they would initiate in their teaching if their students didn't have to take TRACE at the end of the year, nearly all participants suggested that there would be more creative and communicative activities and more allocation of class time to speaking tasks. It should be noted here that teachers' response indicated negative washback of TRACE which was characterised as narrowing down the curriculum to tested skills. Teachers also concurred that not only what they teach but also how they teach have changed as a result of this negative washback.

Another interview question surveyed whether teachers did any special preparation for TRACE. All respondents concurred that they utilised exam oriented teaching methodology such as asking students to do practice exams and working on exam strategies. Another frequently mentioned change includes focus on tasks such as creative writing, role plays, drama activities which would foster students' creativity. Many teachers expressed that they would abandon multiple choice questions and ask more open-ended questions in order to engage students into more critical thinking, responding to texts, and interacting with the texts as a reader. Similarly, it was suggested that there would be more tasks which require students to linking ideas to one's own experiences as well as extensive reading and summary writing. Furthermore, some teachers stated that they would shift the focus on vocabulary other than academic wordlist and spend time on more productive activities for vocabulary. Many teachers stated that they would make use of more communicative, productive student centred activities taking into account students' preferences. In other words, students would have been provided with opportunities to decide on tasks. It was also claimed that there would have been more variety in terms of tasks and materials that would be related to daily life and to students' departments "not just this test".

Finally, teachers were asked about their ideas about how well they think the course supported their students to learn English and be successful on TRACE. All respondents claimed that the preparatory program (especially for the Advanced level students) supported students to a great extent in language and skills that were necessary to pass TRACE. There were remarks

about close alignment of course content and the exam content such as the following: *"I think the course helps students greatly because we do not expect them to do something really different from what we have done in classroom. Note-taking and answering the questions, I think yes, everything is parallel with what we did in class and TRACE."*

However, some teachers voiced their belief that the program was beneficial in passing the test but not learning 'real English" by saying "we don't teach real language" and "Learning English no, TRACE yes".

> *When I think about our system, our aim is to teach students English, right? But, during the modules, it changes, because students are concentrated on the exams, they want to pass the module. So, in a way we start to give exam preparation. When we think about our supplementary materials, our course-books- they are also supplemented- they are multiple choice questions. But, if you are teaching someone English and if your aim is to really teach English you should not have lots of multiple choice questions, because it not real teaching. They need to understand the content; they shouldn't choose the answers from the options. Because in real life, it is not like this. Nobody gives you the options after asking something. So, I don't know...we need to change something in the system I think.*

4.2.8. Findings of classroom observations for washback

Examining classroom activity quantitatively, it was found that the focus was on the frequency of 1) activities focusing on the language skills or knowledge, 2) the amount of time allocated to teacher and student talk, and 3) the students' work mode during the activity. Data are summarised in the following table.

Table 44

Summary of class observations

	No of students	Skill/ Knowledge Focus	Student Modality	Teacher talk time	Student talk time
Upper Teachers					
T1	15	Listening, Vocabulary	Group, Individual, Choral	19',11"	13',54"
T2	15	Listening	Individual, Choral	31',30"	3',25"
T3	14	Listening	Individual, Choral	29',40"	5',22"

T4	15	Integrated	Pair, Individual	22',38"	9',15"
T5	13	Listening	Individual, Choral	27',06"	5',40"
T6	14	Reading	Individual, Choral	26',15"	6',32"
T7	11	Reading	Individual, Choral	32',56"	6',38"
T8	11	Listening, Grammar	Group, Individual	34',55"	4',10"
Advanced Teachers					
T9	14	Integrated	Individual, Pair, Choral	22',58"	4'
T10	12	Integrated	Individual, Pair	28',18"	8',08"
T11	15	Writing	Individual, Pair	3'3	3',35"
T12	14	Integrated	Individual, Pair	31',2"	4',32"
T13	11	Reading	Individual, Group, Choral	24',22"	6',15"
T14	14	Reading	Individual, Group	25',15"	9',05"
T15	12	Listening	Choral	32',26"	2',15"

In most of the observed classes, the skill focused on was listening followed by reading. Except for one teacher, all teachers were observed to have been using supplementary materials whose content were aligned to TRACE content through text length, genre, and exam type multiple choice questions. Majority of the teachers (73%) seemed to have adopted skills based approach in their instruction depending on the activities in the supplementary materials strictly, without integrating other activities and/or sources to practice four skills in a more evenly approach. In contrast, the remaining 27% tended to make use of a variety of other sources (e.g. short videos, texts from internet and other books and photographs) and taught with an integrated skills oriented approach.

Also, data revealed variety across classrooms regarding students' work modality during instruction. Although some teachers were inclined to ask students to work in pairs and groups, the majority opted for individual working mode through requesting students to work on their own. They also tended to ask individual students to respond to questions. There were few instances of group work. However, working alone and responding to teachers' questions in a choral mode were the most frequently observed student activity in class. The interactional pattern in most of the classrooms followed the traditional Teacher Initiation-Learner Response-Teacher Follow Up (IRF) communication pattern (Sinclair and Coulthard, 1975 in Huang 2009). Also, the classroom interaction had a great tendency to be exclusively between the

150

teacher and the students whereas little classroom communication took place between students. Short, often one or two word responses in most of the classes displayed in the student turns in the timed analysis can be regarded as indicators of limited allocation of class time to communicative activities. Questions addressed to students by their teachers stemmed from course book or the general topic of the supplementary material and did not lead to a communicative situation in which students can improve their communicative skills.

As for time allotment to teacher talk and students talk, it was observed that there was a substantial variation between time allocated to teacher talking time and student talking time. The former was considerably higher than the latter. As it can be seen in Figure 2 below, on the average, teacher talk took up 28', 21" in upper-intermediate level classrooms and 28', 10" in Advanced Level classrooms of the total class time of 50 minutes. This variation can be attributed to activities carried out in the lessons because the skill focus of the lessons was never mainly speaking in any of the classes observed. The classes observed were mainly devoted to skills that appear in the exam. This also accounts for the low tendency for the communicative aspect in student to student interaction pattern in classroom communication. Therefore, as previously stated, the elements of classroom communication corresponded to a negative washback effect of TRACE because data extracted reiterated what got tested got taught.
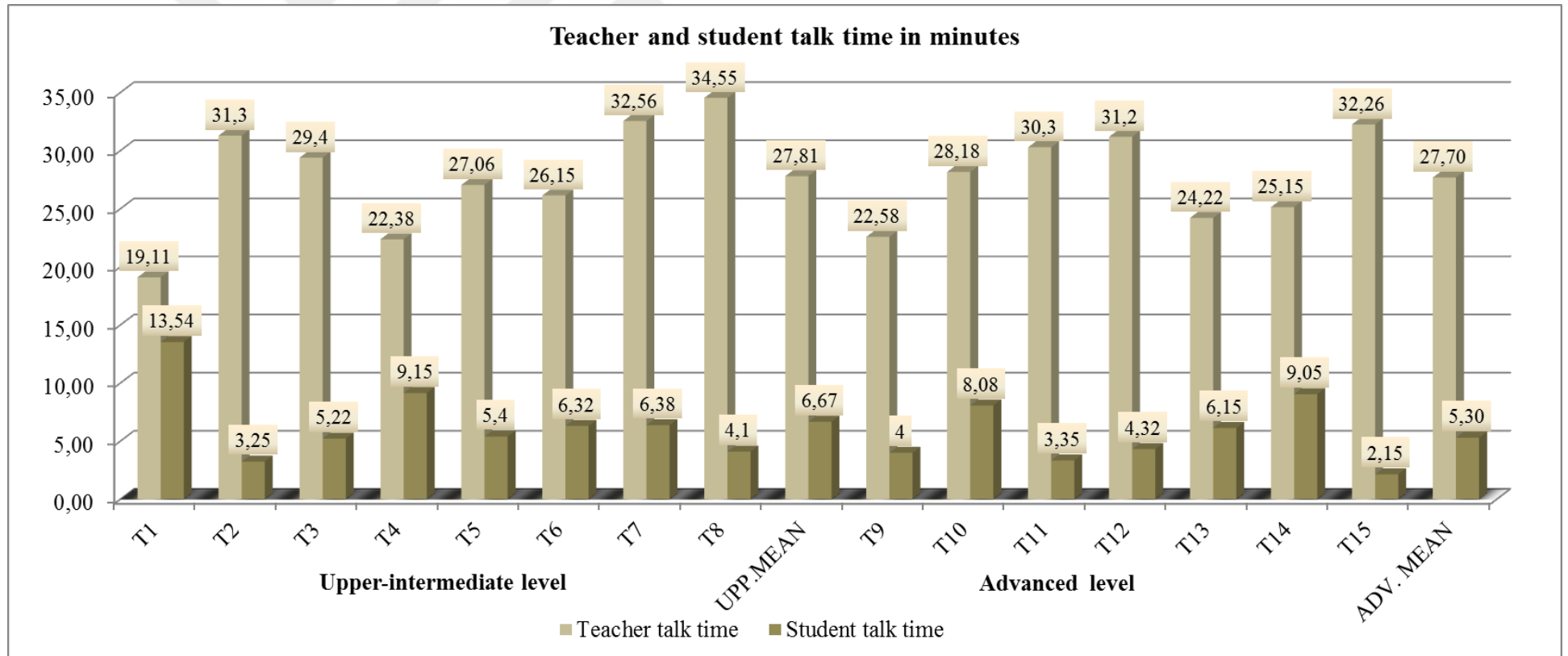
Figure 2: Bar graph showing time allotment to teacher and student talk in observed upper-intermediate and advanced level classrooms

Observations revealed some marked differences among teachers in terms of instruction in four areas: the exploitation of materials, focus on integrated skills and discourse synthesis, explicit strategy training and exam coaching.

Exploitation of Materials

It was evident in the observations that nearly all teachers employed supplementary materials in their classes and some of the teachers displayed great variety in terms of exploitation of teaching materials in comparison to others who covered the materials verbatim. To illustrate, some teachers were observed to raise awareness on critical thinking and reading strategies by replacing the exam type multiple choice questions with open-ended ones projected on power point slides. Teachers expressed that they wanted to raise awareness on the importance of critical reading skills through responding to open-ended questions though some students resisted such type of questions claiming that these questions did not mirror exam type items. Some teachers scaffolded the given supplementary materials/ worksheets. As an illustration, one of the teachers chose to first ask students to read and find the main idea and discuss it in groups. Afterwards, they were provided with the questions of the supplementary materials which consisted of exam type multiple choice items. The teacher required them to match the paragraphs with given main ideas as a follow up activity of their previous discussion. Similarly, although summarising and paraphrasing were not directly tested by TRACE, a few of the teachers required their students to summarise paragraphs in one sentence as an alternative to multiple choice comprehension questions.

In addition, it was observed that some teachers were very vigilant in observing students' topic familiarity and bridging the gap between students' existing and expected background knowledge in reading materials. To illustrate, in one of the observations, the topic of the reading

lessons was "Cubist Art". The teacher, anticipating deficiency in students' background knowledge, made use of extra sources about the topic to increase their topic familiarity.

Observations also unveiled disparities with regard to the use of multimedia through multi layered integrated materials. Although the majority of the teachers relied on utilizing the supplementary materials without modifications, some teachers resorted to integration of extra oral and written sources. Students were required to work across different resources and practice different language skills. For example, one of the teachers used an authentic video and asked students to listen to it and answer specific questions as a lead in to a reading supplementary material. It was observed that some teachers displayed a deliberate effort towards integrating skills through use of authentic materials such as videos and texts from newspapers and blogs. It was also pointed out by some teachers that in TRACE students would be required to read and listen to different sources and make use of these in the writing section of the exam. It can be inferred that some teachers had a succinct understanding of TRACE specifications and adapted the theme-based integrated nature of the exam into their methodology. Based on classroom observations it was concluded that some teachers drew on a more restricted range of teaching materials whereas others supplemented the materials at hand with a variety of other resources.

Focus on integrated skills and discourse synthesis

Prior to observations, data extracted from teacher questionnaires on washback revealed that teachers claimed to have a solid understanding of the requirements of TRACE. However, during observations, except for one instance, it was observed that a deliberate focus on integrated skills approach and discourse synthesis was completely lacking. Very few teachers seemed to have adopted an integrated skills approach in their methodology. Actually, it seemed that instruction depended on supplementary materials on a large scale because most teachers followed the activities in the supplementary materials strictly without making modifications. To illustrate, one of the teachers made use of a short listening as a lead-in to the reading text which was the

154

focus of the lesson. Afterwards, the teacher required the students to refer to the ideas they were exposed to in listening and reading in a follow up speaking activity. That is, some teachers were able to adapt teaching materials and have students practice integrated skills. Even if the major teaching objective of the lesson was to focus on a certain skill, teaching in some classes included a broader range of skills and displayed an even coverage of four skills though lead-in and follow up tasks. Such classes were observed to be less teacher centred since more classroom time was allocated to activities that required students to engage in speaking and writing.

At the onset of the study it was assumed that implementation of a theme-based and thematically integrated proficiency test modelling real language use will affect classroom instruction and bring forth an even focus on teaching of skills. When these results are examined in view of the initially assumed washback effect of TRACE, it can be inferred that there are varying degrees of washback exerted upon individual teachers.

Explicit strategy training

In some classes, covering the assigned teaching material was prioritised, but in some others mastering the focal points of the lesson was attempted through explicit skills training. To illustrate, in one of the classes after the lead-in to a reading task, the teacher asked about the number of reading texts in TRACE, length of these texts, questions types and time allotment. After a brief discussion, the teacher handed out the reading text and asked the students about the time it would take them to read and understand the text. Then she asked whether it would take them shorter to read the text if she also gave out the multiple-choice reading questions. Students stated that it would take considerably shorter for them to read if she had given the reading questions because they would only read the related parts not the full text. Then the teacher distributed a small rectangular printed paper to each student and asked them to punch a small hole in the middle. She said that it was a new reading gadget which would enable them to read better and asked them to read the text by placing the small rectangular paper on the

155

reading text and read through the pinhole that they had punched. After experiencing the difficulty of focusing on isolated words that were visible through the hole in the middle of the rectangular paper, students concurred that reading without understanding the meaning relations between words and sentences was quite difficult. They discussed the similarities of this experience with reading a text only for the specific answers without grasping the main ideas and suggested ways to read efficiently. In addition, it was observed that some teachers raised awareness on skills training by engaging students in activities such as examining title, heading and visual clues to guess the content of texts, exploring main ideas (examining key words, analysing examples given and asking questions), looking for specific information and where it is usually located in the text, and having students discuss the strategies they employed for reading more efficiently. On the other hand, in some classes, teachers read the texts paragraph by paragraph with the students, directing students' attention to context clues to understand the main ideas when they encountered difficulty in decoding the main ideas.

It was also observed that some teachers explicitly taught students how to recycle and recall vocabulary through modelling, working on word forms and vocabulary games. Similarly, one of the teachers focused on note-taking skills strategy training by asking students to compare notes that they had taken with those of their peers and elicited techniques that could be used for efficient note-taking.

In sum observations revealed that while some teachers focused on explicit note-taking, vocabulary and reading strategy training, some others just set the task and then shared the answer key with their students. In the former the teacher highlighted the learning process where as in the latter the focus seemed to have shifted to product.

Exam coaching

Many teachers especially in the advanced level were observed to utilize test-related activities. Some of these activities include drawing a link between class activities and the exam,

156

raising students' attention to the tasks by referring to the exam, giving the students tasks under exam conditions, evaluating students' performance in comparison to exam related task, and pointing out exam taking skills and strategies.

4.2.9. Findings of TRACE scores on different administrations for instructional sensitivity

TRACE as a pre-test was administered in September 2014 to a total of 813 test-takers. 147 of these students were placed in upper-intermediate and advanced levels in PEP. After one term of instruction, TRACE was administered as the post-test in January 2015 to a total number of 324 students including the 147 participants of this study. Both tests included 40 reading items intended to measure students' ability to identify main idea, specific information, inference, text cohesion at the word and sentence levels, guessing meaning of the words from context, and responding to cross-textual questions. The test also included 25 listening items which accompanied lecture listening and performing while listening (responding to comprehension questions while listening to a text) tasks. The overall reliability of the pre-test was 0,60 and the overall reliability of the post-test was 0,62 (using Cronbach's α). These reliability indexes were not high enough for a high stakes test like TRACE. Further, the reading section of the test also yielded low reliability with the alpha level of 0, 54. Reliability of the listening section of the test was even lower (α = .40). Unfortunately, the post-test also yielded low reliability for the reading (α = .50) and listening sections (α = .44). To gain insight into possible relationships between understanding of overall scores and other variables such as reading scores, listening scores and writing scores a correlation analysis was carried out.

Table 45

Correlation analysis of scores

		Reading September	Listening September	Writing September	Overall Grade September
Reading September	Pearson Correlation	1	,25[**]	-,22[**]	,55[**]
	Sig. (2-tailed)		,00	,00	,00
Listening September	Pearson Correlation	,25[**]	1	-,07	,69[**]
	Sig. (2-tailed)	,00		,38	,00
Writing September	Pearson Correlation	-,22[**]	-,07	1	,46[**]
	Sig. (2-tailed)	,01	,38		,00
Reading January	Pearson Correlation	,25[**]	,14	-,20[*]	,09
	Sig. (2-tailed)	,00	,09	,02	,27
Listening January	Pearson Correlation	,12	,13	-,13	,06
	Sig. (2-tailed)	,14	,13	,12	,46
Writing January	Pearson Correlation	-,15	-,14	,19[*]	-,05
	Sig. (2-tailed)	,08	,08	,02	,57
OverallGradeJan	Pearson Correlation	,06	,02	-,02	,04
	Sig. (2-tailed)	,46	,80	,83	,68

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

These statistical values that are of significance since they indicate low correlations between constructs of these two different proficiency tests. Some reasons for low correlations might have stemmed from restricted range of the scores and unsystematic performance on the test due to unclear test construct. However, there was a significant negative correlation between writing and reading scores of the September test r (145) = -.22, p < .01. The emergence of a negative correlation could be a function of using a sample of truncated data. The test scores were truncated because they belonged to students who failed the proficiency test and remained for instruction in the PEP.

An ANOVA with repeated measures was used to compare mean scores of the students in reading, listening and writing as well as overall scores in TRACE that was administered in September 2014 and in January 2015 where the participants were the same in each group. Students who received instruction at the upper-intermediate and advanced levels were measured multiple times to see whether TRACE was sensitive to instruction.

Data was screened against the assumptions of ANOVA with repeated measures. Shapiro-Wilk test of normality showed no violation of this assumption. (for September test scores S-W= .96, df=147 and for January test scores p=.00 and S-W= .97, df=147, p=.00). ANOVA results showed significant differences between pre and post test scores as an indication of instructional sensitivity of TRACE (reading comprehension $F(1, 146) = 234.90$, p =.00, listening $F(1, 146) = 966.88$, p =.00, and writing $F(1, 146) = 264.25$, p =.00). The results demonstrated that there was a significant effect of instruction on overall gain scores, $F(1, 146) = 969,45$, p= ,000. Results also demonstrated that there was a statistically significant increase in scores between the pre- and post- test as a function of instruction. Mean differences between reading, listening, writing and overall scores from pre- to post- test are outlined in Table 46 below.

Table 46

Mean differences between reading, listening, writing and overall scores from pretest to posttest

Measures	Pre-test		Post-test	
	Mean	*SD*	*Mean*	*SD*
Reading Comprehension	18,36	3,11	22,81	2,59
Listening Comprehension	14,57	3,45	25,14	2,73
Writing pretest	19,52	3,58	26,01	4,00
Overall Grade	52,45	5,74	73,97	6,30

4. 3. Findings related to consequential validity

4. 3 .1. Findings of teacher questionnaire for viability of test decisions

17 university instructors of mainstream courses teaching at architecture, psychology, engineering, hotel management, gastronomy, business administration departments were surveyed through a questionnaire. The first question inquired about strengths and weaknesses of students who completed PEP regarding their English proficiency and academic skills. The table below outlines their response.

Table 47

Perceived strengths and weaknesses of the students who completed PEP

	Strengths	f	%	Weaknesses	f	%
Writing	Essay organisation	12	71	Poor grammar	11	59
	Taking course notes	1	6	Lack of vocabulary	5	29
	Taking reading notes	1	6	Citation skills	5	29
				Critical thinking	4	24
				Writing an organised essay	3	18
				Flow of ideas	3	18
Reading	Understanding the main ideas	4	24	Critical thinking skills	4	24
	Guessing words	2	12	Inference	3	18
	Scanning	2	12	Vocabulary	3	18
	Reading short texts	1	6	Reading long texts	4	24
	Following text organisation	1	6	Finding main ideas	2	12

		n	%		n	%
	Inference	1	6	Grammar	1	6
	Skimming	1	6			
Listening	Note taking	4	24	Lack of motivation to listen to long lectures	2	12
	Finding the main and the supporting idea(s)	1	6	Difficulty to understand lectures	6	35
	Inferring attitude and purpose	1	6	Difficulty to understand class discussions	1	6
	Identifying signal words	1	6	Difficulty to understand accents	1	6
				Identifying key words	1	6
Speaking	Presentation	3	18	Unwillingness to speak in English	8	47
	Turn-taking in discussions	1	6	Lack of fluency	3	18
	Expressing opinions and communicate with the other students	1	6	Lack of accuracy	2	12
				Pronunciation	1	6
Grammar and vocabulary	Accuracy	3	18	Poor grammar	5	29
				Vocabulary knowledge	4	25

Majority of university teachers (71 %) stated that students were good in organizing their essays regarding making an outline, integrating supporting ideas and writing a thesis statement. However, some teachers (35%) also indicated that they considered essay organisation as a weak area which requires further work especially regarding mastery of flow of ideas which is framed as inability to convey their thoughts in a properly organized academic structure and building up their own arguments. Another weak area spotted involved students' use of grammar and vocabulary. Most of the teachers (59%) noted the inefficiency of using accurate grammar and lack of adequate vocabulary knowledge (29%) as a major handicap. One comment stated "students sometimes stock phrases and collocations that are wrong. They complete their work with a limited number of words: Therefore, written assignments generally look so simple and lack depth of adequate discussion". Teachers further noted that students' use of grammar was inaccurate due to lack of "confidence in constructing sentences and/or employing alternative

structures", incorrect use and fossilised errors in certain structures such as passives, writing complex sentences, punctuation, and using formal language. Another weakness in writing skill was regarded to be associated with inadequate citation skills pertaining to paraphrasing, quoting and making use of APA by 29% of the respondents. In addition, 24% thought that the students' critical thinking skills manifested through answering questions requiring analytical thinking, editing one's own work as well as peers and integrating research findings into their own work were lagging behind.

In terms of students' reading skills two of the respondents (12%) indicated that students were good in guessing vocabulary from context and good in scanning because they had the ability to read fast and find right answers to the multiple choice questions. On the other hand, it is the contention of majority of the teachers (60%) that students lacked advanced reading skills which required inferencing-identifying tone and purpose, (18%) drawing conclusions through critical thinking, analysis and synthesis of main ideas (24%), coping with comprehension of long texts (24%) and finding main ideas (12%). In addition, 18% of the teachers noted difficulty with guessing words in context since students lacked "awareness of different shades of meaning".

Regarding students' strengths and weaknesses in listening strengths were listed as note taking (24%), finding the main and the supporting idea(s) (6%), inferring attitude and purpose (6%) and identifying signal words (6%). On the other hand, some teachers pinpointed negative perceptions related to students' difficulty to understand lectures (35%), lack of motivation to listen to long lectures (%12), difficulty to understand class discussions (6%) and difficulty to understand accents (6%).

As for speaking skills, there were few comments indicating students' strengths. These were framed as giving presentation (18%), turn-taking in discussions (6%) and expressing opinions and communicate with the other students (6%). Whereas, majority (77%) denoted students' inadequate speaking skills. According to university teachers, students were unwilling

to speak in English (47%) and their speech lacked fluency (18%) and accuracy (12%) as well as correct pronunciation (6%). Speaking skill was defined as 'the weakest point of an average student'. "They feel so insecure when they speak in English. I assume it is due to the lack of confidence in speaking a foreign language". Therefore, it was suggested that students needed more instruction and practice in speaking in English.

The second question on the questionnaire surveyed teachers' perceptions regarding how well students who completed PEP were prepared regarding their English language ability and academic skills. They were asked to mark their opinion on a 4-point Likert scale and their response is listed in the following table.

Table 48

Frequency of teacher perceptions regarding how well prepared PEP graduates in English and academic skills

	not prepared		fairly prepared		prepared		well prepared	
	f	%	f	%	f	%	f	%
Reading academic texts and understanding the main ideas	2	12	9	53	5	29	0	0
Taking reading notes	1	6	13	77	2	12	0	0
Understanding lectures	1	6	8	47	7	41	0	0
Taking listening notes	3	18	8	47	4	24	1	6
Writing an organised essay	3	18	9	53	3	18	0	0
Discussing ideas and expressing opinions clearly and accurately in their speech	9	53	5	29	2	12	1	6
Asking questions	8	47	5	29	3	18	1	6
Using a range of vocabulary appropriately	5	29	11	65	1	6	0	0
Using a range of grammatical structures in their written and spoken work	7	41	9	53	1	6	0	0
Using different sources (notes, summaries etc.) to support ideas in their written and spoken work	10	59	4	24	2	12	0	0
Giving feedback to peers	7	41	4	41	2	12	0	0
Revising own written work based on given feedback	5	29	7	41	3	18	0	0

Some respondents expressed negative perceptions regarding reading for the main ideas by indicating students were not prepared (12%) and some claimed students were fairly prepared (53%). As for taking reading notes students were regarded as not prepared (%6) and fairly prepared (77%). Only two of the respondents (12%) considered students 'prepared' in reading note-taking skills. As for listening skills some respondents indicated positive perceptions stating that students were prepared in understanding lectures (41%) whereas for %53 of the students claimed that they needed to improve this skill. Similarly, for %65 of the teachers, students were not prepared in taking listening notes. On the other hand, 35% of the teacher considered prep program equipped students with note-taking skill.

In addition, the questionnaire elicited suggestions of university teachers to improve the PEP. These point addressed were improvement of speaking skills (35%) and writing skills (29%) as well as more focus on teaching grammar (6%), critical thinking skills (12%), summarising, paraphrasing and citing (12%). Overall, improving speaking skills was the most commonly stated suggestion by the participating teachers. They stressed on the student need to practice speaking much more and gain confidence. One of the respondents noted:

> *I believe in the rigor of the Preparatory School Program at Ozyegin University, however, faculty members share a common belief that students have a big problem in speaking. I would strongly suggest the PSP to encourage more speaking in their classes and to test students' speaking ability maybe with a different method.*

Here it is also important to focus on the suggested idea of "testing speaking ability" in order to improve this skill since it signifies the reliance on testing as a lever for change.

Another respondent suggested: *"Some students can pass the proficiency exam despite low writing skills as they get higher grades from the other parts of the exam. There should be a limit in writing to be able to study at their departments. For example, they have to get a minimum score of 60 from the writing section in TRACE to be able to take ENG 101 writing"*. It can be inferred that some teachers believe that students can be successful on TRACE due to being test-smart. Therefore, taking assessment-driven measures, such as focusing on and

prioritising writing, is seen as an effective way of maintaining higher language competency level.

Another suggestion concerned working on skills that would facilitate skills such as summarising, paraphrasing and basic citation in APA style. One comment concurred: "In upper-intermediate and advanced level students should learn how to summarise articles/videos and write response paragraphs. They have difficulty in summarising and reflecting on sources both in writing and speaking".

In addition, necessity of a deliberate effort towards teaching critical thinking skills through integration of research skills and open-ended questions to test design was mentioned by two of the teachers (12%). It is important to note here that one comment associated the test design with washback and stated that test design, having open-ended questions in this case, might lead to positive test effect leading to better learning outcomes regarding critical thinking. *"Although the current system, I mean assessment, does not allow it to require open-ended questions during the Mid-Module Achievement Test and Level Achievement Test (LAT) in English Preparatory Program and even in TRACE, there must be ways to integrate this into the present practices as it greatly helps students develop their critical thinking skills".*

Another suggestion was related to placing a greater emphasis on structure and form to improving written fluency by *"encouraging a variety of sentence structures, encouraging constructing complex sentences, greater care taken in avoiding habit forming clichés and collocations and stock phrasing that in many cases are grammatically wrong".*

In sum findings of the teacher questionnaire for consequential validity revealed that teachers claimed freshman students who completed PEP had certain weaknesses in terms of grammar competency in English, listening and comprehending lectures and speaking skills. Majority of the teachers (79%) suggested that students had inadequate speaking skills. It was indicated that some (53%) held the belief that PEP did not prepare students in speaking skills

(e.g. discussing ideas and expressing opinions). In addition, according to 59% of the teachers, students lacked skills related to discourse synthesis (e.g. using different sources to support ideas in writing and speaking. These findings regarding perceived weaknesses do not agree with students' perceptions which are explained in the next section.

4.3.2 Freshman questionnaire for consequential validity over time

In an attempt to examine viability of TRACE based test decisions over time, a questionnaire was given to 39 freshman students who received instruction at PEP and had completed upper-intermediate and/or advanced levels.

The first question on the questionnaire gathered information on students' departments. Further information regarding the majors of participating students are outlined in the table below.

Table 49

Departments represented in the consequential validity questionnaire

Department	Number of respondents
Architecture:	4
Interior Design	4
Management Info Systems:	1
Industrial Engineering	6
Computer Science	3
Mechanical Engineering	3
Hotel Management	5
Civil Engineer	1
Psychology	1
International Relations	1
Economics	1
Entrepreneurship	1
Business Management	1
Banking and Financial Management	2
International Business and Trade	2
Business Administration	1
Gastronomy	2

As a response to the second question, students were instructed to check the degree on a 4-point Likert scale in the table on the questionnaire that reflected their point of view regarding how well they think the preparatory program prepared the students in a range of skills.

Table 50

Frequency of student perceptions regarding how well prepared PEP graduates in English and academic skills

	not prepared		fairly prepared		prepared		well prepared	
	f	%	f	%	f	%	f	%
Reading academic texts and understanding the main ideas	1	3	7	18	20	51	10	26
Taking reading notes	2	5	11	28	18	46	7	18
Understanding lectures	1	3	9	23	20	51	9	23
Taking listening notes	1	3	12	31	12	31	14	36
Writing an organised essay	3	8	6	15	13	33	17	44
Discussing ideas and expressing opinions clearly and accurately in their speech	2	5	14	36	18	46	5	13
Asking questions	2	5	10	26	19	49	8	21
Using a range of vocabulary appropriately	4	10	12	31	16	41	7	18
Using a range of grammatical structures in their written and spoken work	4	10	11	28	18	46	6	15
Using different sources (notes, summaries etc.) to support ideas in their written and spoken work	5	13	14	36	15	39	5	13
Giving feedback to peers	3	8	15	39	12	31	9	23
Revising own written work based on given feedback	3	8	8	21	13	33	14	36

Students response revealed that PEP graduates perceived themselves prepared in English and academic skills including reading academic texts and understanding the main ideas, using a range of grammatical structures in written and spoken work, revising own written work based on given feedback understanding lectures, taking listening notes, and writing an organised essay. Overall, it can be inferred that majority of the students suggested positive perceptions in most sub-skills on the contrary of their teachers.

However, in some certain skills such as discussing ideas and expressing opinions clearly and accurately in their speech despite 59 % of the respondents who stated that they were prepared (46 %) and well-prepared (13 %) in expressing opinions and discussing ideas, some respondents claimed that the program did not prepared them (5%) and prepared them fairly (36%). Similarly, in another speaking sub-skill, asking questions, although majority of the students 69% expressed positive perceptions, 30% of the respondents stated that they didn't feel themselves prepared in this skill. With respect to using a range of vocabulary appropriately, there were a variety of responses. For 13 % course did not prepare them. For 36% it fairly prepared them. On the other hand, 37% expressed that program provided them enough support in using a range of vocabulary appropriately and 13% stated that it prepared them well in vocabulary usage.

Students' response displayed diversity regarding how well the PEP assisted them with regard to using different sources such as notes and summaries to support ideas in their written and spoken work. 49% of the participants responded negatively implying that they did not feel fully equipped in synthesizing information across sources in their own work. On the other hand, 52% expresses that they felt prepared to utilize different sources as a means of developing their arguments in their written and spoken work. Similarly, the question related to 'giving feedback to peers' yielded both negative and positive response.

When freshman students were asked to profile the strengths and weaknesses of the PEP graduates, they commented on the following.

Table 51

Perceived strengths and weaknesses of the freshman students who completed PEP

	Strengths	f	%	Weaknesses	f	%
Writing	Writing an academic essay	18	46	Using a range of grammar structures	3	8
	Using grammar	1	3	Vocabulary knowledge	6	15
	Using vocabulary	1	3	Academic writing	4	10

Skill	Item			Item		
				Punctuation	1	3
Reading	General comprehension	9	23	Lack of vocabulary knowledge	2	5
	Critical reading (Inferencing and questioning)	1	3	Critical reading	1	3
				Reading for main ideas	1	3
	Vocabulary	1	3	Concentrating on long texts	1	3
				Summarising	1	3
				Understanding supporting ideas	1	3
Listening	Note-taking	7		Note-taking	2	5
	Understanding the main ideas	5	13	Understanding main ideas	1	3
				Difficulty to understand different accents	1	3
				Lack of vocabulary	1	3
Speaking	Doing presentations	1	3	Speaking fluently	9	32
	Using accurate grammar structures while talking	1	3	Using newly learned structures	1	3
				Pronunciation	1	3
	Asking-answering questions	1	3	Using newly learned vocabulary	1	3
				Expressing own opinion	1	3
Grammar and vocabulary	Vocabulary expansion	3	8	Using grammar accurately Practice	1	3
	Using grammar accurately	4	10	Using vocabulary in writing	1	3
	Using vocabulary in writing	2				

Finally, freshman students offered a variety of measures for improving the PEP including having more practice in grammar and vocabulary and all skills but especially in speaking, learning about basic citation skills as well as APA, and integrating an EAP flavour especially by learning vocabulary related to departments. One respondent stated it was necessary to add a speaking component to the proficiency exam, TRACE. One of the comments indicated *"passing course is easy, needs to be more difficult"* and another stressed *"Advanced level is more simple than upper. It should be more difficult"*. Also there were some comments which highlighted the concept of variation between teachers in the PEP by stating: *"PEP needs to self-check about the teachers and the application of the plan (curriculum). Plan and the opinion of education is okay but some problems happen on the stage"*.

4.3.3 Interviews on viability of TRACE-based test decisions with the university instructors of

mainstream courses

Teachers were surveyed regarding their opinions about the language skills required to

succeed when students move on to their departments after PEP. Their perceptions are

presented in the table below.

Table 52

*Teacher perceptions of language and academic skills required to be successful at the
department*

Language and academic skills required to be successful at the department	f	%
Reading	16	84
Listening	10	53
Writing	12	63
Speaking	18	95
Grammar	3	16
Vocabulary	7	37
Academic Skills (Discourse synthesis, knowledge of APA, citing, evaluating online sources, critical thinking)	10	53

Speaking (95%) and reading skills were mentioned the most frequently (84%) by the

teachers to achieve academic success at different departments. In addition, teachers stressed

that some of the speaking skills that are required at the departments involved giving oral

summaries (5%), responding to texts orally (11%), carrying out presentations (32%),

participating in group discussions (26%), asking and answering questions (100%). In terms of

expected writing skills, teachers referred to summary writing (16%), paraphrasing (16%), citing

(32%), writing an argumentative essay as a research follow up (5%), response writing (16%),

writing reports (16%), and projects (21%).

In the next part of the interviews, teachers were asked whether they think students who

completed the PEP can cope with the academic demands of their departments. 2 of the

informants responded positively (10,5%) by stating that PEP was offering students

opportunities to learn English and it depended on the students to make the most out of these

learning opportunities. One of the participants remarked the role of students' taking responsibility of their own learning by saying the following:

> *You clearly see that it depends on the student and how he or she takes it seriously. Today I had a presentation, students were presenting and I know these students from prep program. Two students…they were gorgeous. They did everything, that's a professional should be doing while presenting. You can see that students are hardworking and they took all the things in prep seriously and they benefitted from each and every bit in prep. If the student is hard working, he comes with developed skills and we try to build up on it. But with the students who take things for granted, we have issues with the language. Still you can see expressions like I'm agreeing with or she is come or he is come, the basic grammar mistakes and fossilized mistakes. But this is I think because the students didn't take courses seriously and they do not put that much effort.*

On the other hand, 7 teachers (37 %) opted for a negative response. They gave the following reasons: (1) time span of one academic year allocated to learning a foreign language was not enough, (2) students lack long term goals of learning a language, (3) students had great difficulty in speaking English, (4) students lack skills into how to learn and (5) low level of general English competency of the students. For 10 teachers (53%), students who completed the PEP can cope with the academic demands of their departments to some extent. It should be noted here that teachers often stressed their expectation was not to observe error free grammar and extensive knowledge of vocabulary. However, they expected to see self-expression and communication of ideas not only in writing but also in speaking. They commented that prep program often equipped students with some writing skills but couldn't cater for speaking skills. One of the respondents commented that there were some students in the departmental courses who managed to complete the PEP and pass TRACE despite their low level of language competency.

> *When I got a class in the very beginning of the semester, I asked how many students came from prep. About twenty out of thirty raised hands. Out of those twenty, five or six will be very good, very well-equipped. Although they are still, especially in speaking they would be very shy and very insufficient. Let's say five out of twenty would be equipped in terms of writing and can do the work in discipline. About ten will not be up to standard, really. So, they struggle. About five, they shouldn't be there at all. And I am trying to be realistic. I mean knowing the context, knowing*

the educational background and the possibilities what could be done in prep school in a certain amount of time, feasibility…. I'm taking in all those factors and I'm trying to give a kind of realistic and generous answer; Thirty per cent of prep school graduates shouldn't be there. They are not ready. They are effectively, really, still in intermediate or even pre- intermediate level, in some cases. And somehow, they managed to slip through the net.

Teachers were also asked to share their comments and observations regarding their students' achievement in writing, reading, listening, speaking, use of grammar and vocabulary as well as other academic skills in order to examine the correspondence between the exit criteria of PEP (TRACE) and expected academic requirements at the departments. Teachers stated that they observed certain weaknesses in students' language and skills. The table below portrays their opinions.

Table 53

Teacher perceptions of students' weaknesses in language and academic skills

Students' weaknesses in language and academic skills	f	%
Reading	7	37
Listening	4	21
Writing	8	42
Speaking	15	79
Grammar	9	47
Vocabulary	10	53
Understanding exam questions	6	32
Academic Skills (Discourse synthesis, knowledge of APA, citing, evaluation of online sources)	3	16

Low level of language competency was deemed as one of the weak areas of students. 47% of the teachers suggested that students had inadequate grammar knowledge and 53% expressed that students had insufficient vocabulary knowledge. Some teachers stated that students' low level of grammar and vocabulary knowledge intervened in their understanding. These teachers stated that they had difficulty while marking students' written tasks since they could not decide whether to take quality of language or content into consideration. This was defined as the "big dilemma".

I gave an assignment to my students and I started giving feedback with grammar instead of focusing on content. I wrote many notes saying; 'did you mean this here?' Then after a while I wrote notes on the papers saying I am tired of correcting errors so please deal with the grammar errors on your own before submitting it.

Consequently, many teachers concurred that they compromised and ignored the language and focused on the content.

Spoken English is generally fine but written English needs a lot more work. Papers are not acceptable. If I was grading, taking full consideration of writing, most of the papers will be F. They have a very big problem. They have very big problems with structure. In vocabulary, they are not competent in use of prepositions. So, generally, speaking, I'm not very happy with that and I don't have time sit down and teach them how to write grammar English.

However, there were comments that language errors were so frequent and serious that they interfered in understanding the text.

You don't understand but you try to guess. So, you try to guess what they actually mean. Most of my friends, we're trying to not to be so harsh. So, okay language is not our main concern, because we're teaching the concept, not language. So, we're expecting, them to understand physics. So, okay, we're not dropping any mark for this, language problem. But we try to understand what they mean and it is difficult.

Some comments drew attention to the necessity of language use in authentic context. They expected operationalizing of the theoretical grammar knowledge into practice in order to raise awareness and assist students better when students encounter uses of these structures in their departmental courses.

I'm guessing that grammar, for example, is got through in preparatory school without probably, I'm guessing, without any kind of anticipation of how that grammar could be used in academic skills. One, two areas that stick out the most are the use of passives, for example, passive has always been problem for our students, and punctuation. They are told in the language classes, probably told as a kind of grammatical form that's useful that there is an option to use. But, it is probably not highlighted enough that is if that's in the utmost importance in the academic writing and the research.

For 42% of the informants, writing skills proved to be both daunting and difficult for the students with respect to content generation, self-expression, citing skills and academic formatting.

I think they are not really good at academic writing. Yes, there are some examples where there are a lot of spelling problems, grammar problems, mistakes, but I have seen a lot of papers that had good grammar and good spelling yet not very good at communicating what they have in mind.

Teachers' response also highlighted some areas in which students' performance was acceptable. Table below presents the teachers' perceptions of students' strengths in language and academic skills.

Table 54

Teacher perceptions of students' strengths in language and academic skills

Observed student strengths	f	%
Reading	1	5
Listening	5	26
Writing	4	21
Grammar	4	21
Vocabulary	4	21

According to 5 participants (26%) students did not have any difficulty when they listened to lectures and videos and videotaped lectures delivered on-line. 21% of the teachers stated that students were performing well in terms of writing, grammar and vocabulary.

Data analysis led to 3 themes which were weakness of speaking skills, effect of educational background on achievement and low motivation for reading.

Table 55

Emerging themes of interviews with department teachers

Themes	f	%
Weakness in speaking skills	15	79
Effect of educational background on achievement		

Exam orientedness	3	16
Mismatch between prior learning and university culture	4	21
Low Motivation for reading	8	42

Lack of speaking skills

Majority of the teachers (79%) conveyed students speaking skills in English.

Unfortunately, most of the students I can say 90%, plus maybe, are unable to follow the class because of the language problem. And, they are unable to ask questions in foreign language. They, they prefer to ask in Turkish. Well, there are foreign students in the classes, right? For example, sometimes they ask questions in order to understand the conceptual things in the class. So, I have to, I have to translate their Turkish questions into English for other students. They don't force themselves and ask questions in English. They cannot ask questions. And that affects the course very bad and negatively. They think that they will not be able to do that and they think that they will be ashamed in class, in front of their peers, in front of their friends. So, they prefer not to ask. And that affects the success rate in the class, at all. Language is a main concern for our studies. We talk, we show, we discuss, we explain and we expect students to interact with us to join to the class, you know, to contribute to the class, ask the questions, discuss the concepts with us. But they do not do anything. They prefer to stay silent and just watch. Then I feel, and most of us feel like, we're just lecturing in front of a wall. That's a big concern"

Despite encouragement from some teachers who stated while speaking in self-expression and fluency were important and the focus was not on grammatical accuracy, most of the students were reported to avoid speaking in English.

Speaking is also important. And I'm okay with a lot of mistakes in speaking, which is fine, which is totally fine. But the problem is sometimes they kind of don't speak at all, because they can't speak English. So what they do is either try to speak in Turkish, which rarely I allow, occasionally I allow; or they don't speak at all, just not speak in English.

Some teachers commented that despite the presence of international students in the class, after a while students are inclined to turn to L1.

If there is an exchange- international student in class, then they try to speak in English but eventually they lock out the English speaking student and start to talk in Turkish. Then the conversation stops. Even if something comes to their mind most of the students do not communicate this. At most time they try to respond to my questions

in Turkish and participate in the discussion in Turkish. They have a great difficulty in self-expression and talking in English. If they come from prep school, they have great difficulty in speaking as well as understanding English but if they come from a good high school they do not have problems in speaking.

The comment highlighted the insufficiency of speaking skills of students who come from PEP. This may be due focusing more on other skills that are tested in the TRACE and ignoring speaking which is not tested.

One of the participants stated*: "There is a general tendency to avoid speaking in English as well as reluctance".* Similarly, another comment highlights students' reluctance to speak in English by reporting the following.

Students insist on doing presentations in Turkish instead of English. I observe a serious insufficiency in students' speaking skills in English. At times there is also lack of self-confidence. Sometimes they know but they don't have self-confidence. So I think they don't have enough practice in speaking.

Another emerging theme concerns the effect of educational background on students' achievement. Some teachers commented that there was a mismatch between previous educational culture which relied on exam oriented approach to learning and route memorisation and university culture which emphasised critical thinking and (re)constructing knowledge by synthesising information from various sources. Some teachers claimed that students do not place emphasis on evaluation of their performance since they are product oriented.

Also, it was suggested that because students had spent quite a lot of time getting ready for high-stake tests in Turkish, exposure to English was minimum and that could lead to a major drawback at the university where the medium of instruction is English.

There is a mismatch between students' educational background and the skills required at the university. Schools are busy with teaching students how to solve a multiple choice question without having the knowledge. How to answer a question without knowing… This is what "dersanes" do. They say we teach the technique. Students focus on the correct answer rather than why that's the correct answer. Often why is never asked. Students say "tell me one thing which would summarise life and be the answer to all questions. Thinking, evaluating, criticising is a mind-set and most students do not seem to have that. Here at the university they need to be formatted and it is very difficult. They come from secondary and high school with

gaps in their knowledge and when they have education in another language they are startled and anxious.

Exam oriented approach to learning was also considered to be an issue related to educational background of students. One of the teachers concurred *"students love multiple choice questions because they use recognition memory.* In addition, another teacher mentioned using an online homework system which multiple choice items and commented: *"Students are very much accustomed to multiple choice items and they prefer responding to this format".*

Another theme that emerged from data analysis was discourse synthesis. Teachers noted that students have difficulty in integrating ideas from other sources into their written work.

> *The ability to synthesize ideas is really wide-challenging. Sometimes, we again, we're not sure whether it is language issue or whether it is a generally critical thinking issue. I mean, synthesizing is putting ideas together. For example, seeing, detecting the patterns, similarities, new connections between them, there is a critical thinking skill. We're not sure, and the students are not generally very good at it. If, I'm honest, I mean, the ones who have gone to private schools, for example, the ones who come straight, you know, by passing prep school, are generally much better, I think. They probably had more exposure or early exposure to this. So, well I'm presuming, grade nine or the grade ten, those, those skills of debate, arguing, putting ideas together, synthesizing views, evaluating whether an argument is good or not. Perhaps that it helps the idea that if those skills are introduced at an early stage.*

Data analysis pointed out lack of reading motivation as another theme. Some teachers acknowledged that teaching a foreign language is a difficult task with in a limited time frame of one academic year as in PEP. Also it was suggested that teaching necessary reading skills to cope with academic requirements in the university program was even more challenging. One of the respondents commented: *"I think 8 months is a very limited time to teach a person who knows nothing about English or a foreign language and then eight months later, giving him a fully academic text about law, about psychology, about industrial engineering. This is a tough job".*

Response of the teachers indicated that students "resist reading". This resistance was at times associated with exam oriented approach that students have towards learning. In other words, it was suggested that students were inclined to respond to a certain type of exam questions which would involve multiple choice format, "*I was really surprised to see that if there is a question which is more than 5-6 lines in the exam they ignore the question and don't do it. They don't even consider to put in the effort to read it. They don't read the instructions to an assignment. They ask and want me to explain. They run away when they see a reading text*".

One of the respondents associated lack of reading motivation in students with the difficulty of synthesizing ideas from different sources. In the following excerpt reading is defined as the bedrock for the discourse synthesis.

> *You need to read so that reading may help you. These students don't read that much. And reading is kind of a burden for them, so if you don't read you cannot practice and the brain cannot get its vitamin in the form of reading. So, integration of these ideas or building argument these may cause trouble.*

Another comment focused on students' lack of motivation in reading and the potential contribution of PEP to students' reading motivation. Suggesting that students aren't encouraged for extensive reading in their prior educational background, some teachers claimed that students have a deep rooted lack of reading motivation not only in their second but also in their native language.

> *I don't know you can do anything in prep language preparation them like apart from of course like you know, you probably assign them social science texts, right? But, if they don't read in life, if they don't even read a novel, in Turkish even, how would they understand what they read, right? So, it is reading in general like, if you don't read general at all, which whichever language you read, whatever language you read, you are not gonna be able to understand what you read, right?*

Even if they are tested in reading in TRACE, this does not exert a powerful thrive and washback to make an effort in acquiring reading skills: *"They don't seem to have understood the logic*

and purpose behind reading skills. They try to memorise and therefore their affective filters are up. They are really anxious".

4.4. Summary

In relation to the first research question, namely, washback of the exam, findings from student and teacher questionnaires and interviews indicated that there may be both positive and negative test effects.

Washback on materials

Questionnaire and interview data unfolded that a large proportion of the respondents agreed course materials catered for improvement of students' competency in skills and language. It should be noted that although speaking was not a tested skill in TRACE students affirmed that materials nurtured speaking skills through a variety of follow up tasks. However, it was suggested by some students that materials lacked exercises for synthesising information from different texts to answer questions, open-ended questions, critical thinking and summarising exercises. In addition, overwhelming majority of the students indicated their belief that writing, reading and listening skills coverage of the materials would help students to achieve better performance in reading on the test. It was also generally agreed that reading and listening texts were followed by different activities but it was also marked that listening materials did not frequently cover 'combining listening notes with other sources to write an essay'.

Findings also conceded that teachers had positive perceptions overall about the course materials and they believed that materials covered skills and language that would help their students to perform well on the test. However, it was also stated that reading, listening and writing materials were unable to cater for summarising, paraphrasing and synthesizing information from different texts to answer questions. Overall, there was positive washback of

179

the TRACE exerted upon the materials but more than half of the teachers pointed out that the above mentioned skills which could facilitate learning were excluded from the materials

Overall, student and teacher perceptions regarding reading, listening and writing materials and how well they nurtured students' improvement in language skills captured in the student questionnaire data revealed that teaching materials and TRACE were closely aligned and this alignment corresponded to the TRACE's positive washback on materials. Students and teachers deemed materials covering tested skills and echoing exam tasks and questions types as a positive contribution to their learning.

Facilitating effect of discourse synthesis

The present research study concluded that reading across sources and synthesizing these ideas into the writing task of the exam may have a facilitating effect upon student learning but it was also remarked that synthesising ideas from reading and listening into wiring can be language level dependent. Many respondents pointed out to the positive washback of the thematically integrated TRACE on learning because it raised students' awareness to generating, organising and linking ideas, topical knowledge as well as modelling vocabulary and sentence structures. However, the study also found that there was a discrepancy between test designers' intended washback and the washback in classrooms due to lack of tasks and exercises which cater for discourse synthesis and varied instruction of individual teachers.

Narrowing of curriculum

Data extracted from student questionnaire about classroom tasks employed by their teachers indicated an even balance of instruction of four skills, focus on synthesizing ideas across different sources and exploitation of communicative activities which foster student to student interactional patterns. Based on students' account the fact that students would take TRACE at the end of instruction did not lead to narrowing of curriculum and omission of

speaking activities since it is not tested in the proficiency exam. However, data also revealed that exam practice and exam taking skills and strategies took up class time on a large scale.

Similarly, this finding is in accordance with teacher questionnaire data which indicated that although exam practice was integrated into teaching and changed what the teachers taught, there wasn't any negative washback of TRACE regarding how teachers taught because similar to students, teachers also agreed that there was practice of 4 skills through some creative and communicative activities. However, there was disagreement regarding use of some collaborative and communicative activities which called for critical thinking and using content and ideas across different sources. Additionally, not only students but also teachers indicated that curriculum was not narrowed down to tested skills such as excluding practice on speaking which was not assessed on TRACE.

On the other hand, based on data canvassed from observation and interviews it can be concluded that the test may have led to narrow curriculum offerings since teachers were employing instructional approaches they would not have used if it wasn't for the exam. Limited variety of classroom organisational patterns, time allotment to tested skills at the expense of communicatively oriented activities and practice of speaking (which is not tested on the exam) supported the finding that negative washback has led to narrowing of curriculum in the classrooms.

Washback on teachers

Teacher and questionnaire data indicated that although exam practice was integrated into teaching and changed what the teachers taught, there wasn't any negative washback of TRACE regarding how teachers taught because similar to students, teachers also perceived that there was practice of 4 skills through some creative and communicative activities. However, there was disagreement regarding use of some collaborative and communicative activities which called for critical thinking and using content and ideas across different sources.

On the other hand, based on student and teacher interview data and classroom observations it can be inferred that for some teachers TRACE may have exerted a negative washback and changed not only what they taught but also how they taught, and affected others at varying degrees since even though they were prepping their students for a high-stakes proficiency exam they did not change how they teach. Consequently, it can be inferred that the TRACE exerted negative washback on teaching because some teachers were more focused on tested skills and were coaching their students to increase their gains in scores instead of deep learning.

Washback on learners

Findings of the interview data of the study pointed out to a negative washback effect on student learning since students were inclined to be test-oriented and perceive activities oriented towards the test or test-specific coaching to prepare for the high-stakes proficiency exam. Several discrepant findings from this study further supported the context-oriented and complex nature of washback (Chen 2002).

As for the second research question, namely how much learning had happened (instructional sensitivity) the results of the ANOVA for repeated measures indicated a significant result regarding instructional sensitivity of TRACE. Data gathered from the student and teacher questionnaire and interviews also supported this finding since they also pointed out that overwhelming majority of the students perceived that they learned a great deal in PEP.

Finally, the results of the third research question investigating the consequential validity indicated that TRACE might have lack of consequential validity.

CHAPTER 5

DISCUSSION AND CONCLUSION

5. 1. Introduction

The aim of this research study was to call the validity of the assumption that employing an integrated theme-based test of English language proficiency which is similar to authentic language use in the tertiary education context in Turkey would bring about positive test influence upon teaching and learning into question. This mix method research study using multiple approaches investigated the washback effect of a thematically integrated proficiency exam on the teaching methodology, materials, classroom activities and learning outcomes of the PEP linked to it as well as consequential validity of the test in a Turkish context in the tertiary education.

In relation to the potential washback of the TRACE, student and teacher questionnaires, interviews and classroom observations indicated that there could be both positive and negative test effects exerted on the teaching methodology, materials, classroom activities and learning outcomes. In addition to positive washback on materials and potential of the test in boosting instructional sensitivity (washback on product) through employment of a theme-based integrated skills approach in testing, this study also found that TRACE could lead to negative washback in the form of teacher and learning strategies geared towards being successful on the test and narrowing of curriculum.

In addition, the proficiency scores obtained before and after the instruction of PEP revealed that TRACE was sensitive to instruction since the competency of the EFL students on the three language skills has highly improved. This finding was also supported by the data obtained through the student and teacher questionnaires and interviews.

Finally, the data obtained through the semi structured interviews with university instructors who offered mainstream departmental courses and questionnaire from freshman students and their teachers reported that the TRACE may lack consequential validity.

The remaining of this chapter first discusses the findings regarding the washback and instructional sensitivity of the test as well as its consequential validity. After that, the pedagogical implications are briefly outlined. Finally, the limitations of the present study are explained followed by suggestions which could constitute the agenda for further research.

5. 2. Washback of the test

The findings of the present study also supported the claim that washback research is complex and multifaceted (Alderson & Wall, 1993; Bailey, 1996; Cheng, 2000, Watanabe, 2004a: Qi, 2004) and corresponded with other washback researchers' contention that investigation of washback phenomenon should include multiple perspectives and triangulation through use of different methods (Bailey, 1999; Cheng, 2001; Wall, 2005, Watanabe, 2004a). For example, Alderson and Wall's observation that "washback is neither simple, nor direct, but circuitous and complicated" has found support in this study (in Qi, 2004, p.188).

Washback on Materials

Findings of the study corresponded with the conclusions of previous research studies that tests may exert positive washback effect on materials (Cheng, 1997; Lam, 1994; Watanabe, 2000).

Discourse synthesis

The body of empirical research to date suggests that the major purpose of washback research is to gather information to guide researchers and educators in engineering positive washback by making the best use of tests. Based on the findings of the present study it is possible to infer that positive washback can be engineered through integrated tests which promote Target Language Use (TLU) through discourse synthesis. Reading across sources and synthesizing these ideas into the writing task of the exam may have a facilitating effect on student learning but it was also remarked that this could be language level dependent. Many respondents pointed out to the positive washback of the thematically integrated TRACE on learning because it raised students' awareness to generating, organising and linking ideas, topical knowledge as well as modelling vocabulary and sentence structures. However, the study also found that there was a discrepancy between test designers' intended washback and the washback in classrooms due to lack of tasks and exercises which cater for discourse synthesis and varied instruction of individual teachers.

Narrowing of curriculum

It has often been claimed that tests lead to 'narrow curricular offerings' and 'what is tested is what gets taught' Findings of the present study are in accordance with the other research studies which found negative washback effect of tests manifested in the form of curricular narrowing through more coaching activities (e.g. focusing on skills and strategies for test taking, analysing questions) and instruction of test-taking strategies (e.g. formulaic approach to teaching writing, less time allotment to untested skills) (Alderson and Wall 1993; Cheng, 2005, Green, 2007b, Hayes and Read, 2003, 2004; Ferman, 2004; Qi, 2004; Wall & Horak, 2006; Azadi & Gholami, 2013).

Washback on teachers

Class observations, teacher and student interviews and questionnaires revealed that the teacher's methodology and choice of classroom activities were adapted to the tested skills and

contents of the test. Teachers stressed that their methodology was affected by the exam because requirements of the exam made them do things they would not otherwise do such as omitting and spending less time on tasks that required speaking. Speaking received little attention since it did not directly contribute to the requirements of the exam. Teachers explained that they had the tendency to employ materials and rely on activities that were compatible with the principles underlying the exam and placed more value on tested skills. These finding are in line with previous research studies which concluded when teachers perceive that teaching and learning are 'circumscribed and controlled' by the examinations and think that students are exam-oriented focus of teaching and learning shifts towards tested skills (Anderson et al., 1990, Widen et al., 1997, Calder, 1990 in Cheng Cheng & Curtis, 2004, p.9: Hayes & Read, 2004)

The present study indicated that both content (what gets taught) and methodology (how teachers teach) may be affected by the exam, but the test may have varying amounts and types of washback depending on the teacher involved. These findings are in line with previous findings of the studies of washback in language education (Alderson & Hamp-Lyons, 1996; Stecher, Chun & Barron, 2004; Andrews, 1994; Blewchamp, 1994; Watanabe, 1996b) and somewhat different from those of previous studies that found no straightforward connection between the test and how teachers taught (e.g. Wall and Alderson, 1993; Chen, 1997).

Washback on learners

Findings of the interview data of the study pointed out to a negative washback effect on student learning since students were inclined to be test-oriented and perceive activities oriented towards the test or test-specific coaching to prepare for the high-stakes proficiency exam. This finding was in line with the contention of other washback studies that students attach importance to skills tested and focus highly on exam-related activities, test content and format (Stoneman, 2006; Shih, 2007, Tsagari, 2009; Gosa, 2004).

The above mentioned results indicated that the test effects on teaching and learning processes, participants and instructional sensitivity (learning outcome) were ascribed to this theme-based and integrated test whose development was informed by the language needs of the learners and the objectives of the institution.

5. 3. Instructional Sensitivity of TRACE

Findings of the study demonstrated that the TRACE was sensitive to instruction. Students of the PEP were able to boost their scores in four months from inception of the program to its completion. However, based on results it can also be inferred that this gains in scores is a function of a variety of factors including teaching towards the test and exam coaching.

5. 4. Consequential validity of TRACE

This study found evidence of positive washback as well as negative washback of the theme-based and integrated English proficiency test, TRACE, that is used as an exit exam in the Preparatory English Program of a tertiary level institution in Turkey. Positive washback following the introduction of a test is regarded as related to consequential validity, whereas, negative washback is associated with lack of validity (Ferman, 2004). Negative washback was mainly exerted on learning and teaching and as a result narrowed the boundary of the curriculum offered and reinforced exam-oriented learning. University mainstream teachers have also confirmed these findings reporting that students who furthered their academic studies beyond PEP have encountered difficulties in their academic competencies and skills due to their insufficient language proficiency. Therefore, the findings of this study point to lack of consequential validity resulting in a narrowing of the curriculum to tested skills which eventually hinders learning.

5.5. Implications

This study demonstrates the importance of employing a combined research framework using multiple approaches to data collection. To this end it resonates with other previous

empirical washback studies (Bailey, 1999; Burrows, 2004; Cheng, 1997, 1998, 2001: Watanabe, 2004). Therefore, one of the implications concerns washback researchers with respect to design of washback research in order to capture the complex and multifaceted nature of the washback phenomenon.

In addition, the findings of the study also revealed significant implications with respect to the instructional design of a Language Preparatory Program as well as teacher development.

Implications for Teachers

Since "teachers are the final arbiters of policy implementation" (Menken, 2008, p.401) teacher perceptions play a significant role upon the motivation and effort s/he makes to improve teaching methodology and affect learning (Clark & Peterson, 1986; Lane, Parke & Stone, 1998 in Pan & Newfields, 2012).Wall and Horak (2008 in Djuric 2008) indicated that there may be times when nature of the tests and intended washback effect might be ambiguous to the teachers and they highlighted the importance of communicating intentions of the testers and underlying principles of tests to teachers. Teachers should familiarise themselves with the exam specifications and be aware of the underlying constructs as well as purpose and the value of the exam. Consequently, they will be able to provide their students with information regarding requirements of the exam and explain the underlying constructs of the exam (e.g discourse synthesis and how this would contribute to their academic studies). Acknowledging the exam requirements would also help teachers modify and adapt classroom materials and exam-preparation materials and integrate learning and testing materials.

Also, teachers should balance teaching and testing in class by avoiding too much focus on exam-coaching and narrowing curriculum down to tested skills. Teachers should devise efficient methods for coping with the build-up of pressure from students to allocate class time majorly to exam preparation activities (mock exams, exam-strategy training, revision, multiple choice items) and focus tightly on tested skills rather than communicative activities. To this

end, teachers could give informative and feedback to diagnose learning needs rather than equating their language proficiency to the number of correct items on a test. Teachers could make use a variety of channels to assessing their students' performance and share their observation with the students.

Another implication for teachers include raising awareness on lifelong learning and shifting focus on students' tendency to adopt the immediate goal of learning English to be successful on the exams to long-term goal of learning a language for communication and other life needs.

Implications for Teacher Trainers

Cheng and Curtis (2004) remarked the importance of interplay and intertwine of influences within each specific educational context where the assessment is put in to use by stating; "the relationship between testing and teaching and learning does appear to be far more complicated and involve much more than just the design of a "good" assessment" (p.16). Increasing assessment literacy of the teachers and raising awareness on constructs of the assessment means would also have an immense effect on teaching and learning. To this end providing teachers with assessment literacy becomes vital. Therefore, teachers should be informed about the constructs of an exam, how to choose and adapt exam preparation materials, how to exploit various instructional methods (including blended learning and use of technology) in dealing with teaching materials.

Implications for Test Designers

Washback studies may play an important role in ensuring ethical language testing by bringing providing opportunities for continuous validation processes and fair test interpretation. The findings of this study may be useful for test designers who wish to engineer positive washback to approach continuous test development effort systemically and maintain an open and transparent communication with the teachers and learners. Test designers can induce

positive washback by providing different stakeholders (e.g. teachers, students…etc.) with information about the purpose(s) and value of the exam, the bedrock attributes (underlying constructs) and scoring procedures rather than general information (e.g. number of items in different sections of the exam). To illustrate, students can be provided with learner friendly versions of the exam rubric and sample items where as teachers can be provided with detailed explanations based on several examples.

Also, test designers have a responsibility to improve the quality of a test. Therefore, test designers should conduct systematic washback studies on a wide array of factors including attitudes and perceptions of different stakeholders alongside exam design and procedures. Furthermore, findings of such research should be disseminated and shared with different stakeholders. Another responsibility of test designers comprises training teachers in becoming assessment literate because tests seem to affect how and what teachers teach. Therefore, test designers should offer teacher training programs to communicate with the teachers, clarifying test constructs and purpose.

5. 6. Limitations

As the research findings are mainly based on analysis of data from a local proficiency test in a specific educational context, it could be argued that generalizability of the findings to the broader English language teaching and testing populations in other contexts would not be possible. However, researchers (e.g. Tsagari, 2006; Perrin, 2000) argued that any washback research is innately context-based. It is imperative to point out here that investigating those forces in a specific educational context may shed a light on similar forces at play in a broader context. The focus of the study was teasing out the washback of a context-specific theme-based integrated proficiency test which mirrors authentic language use in academic settings in tertiary education. Findings can bring about relatability and may have implications for EFL students and teachers and test designers with similar needs in other contexts which aspire to engineer positive washback.

Another limitation relates to one of the data collection instruments; TRACE exam scores. Since reliability of the pre- and post- TRACE exams were low and correlations between sections of the exam revealed low and negative correlations readers should be cautious about findings. However, this limitation does not detract from the value of the present study in being able to explore intricate web of washback mechanism through use of multiple data collection instruments from multiple perspectives.

It should be noted that the analyses would have been more sensitive to relations between proficiency test scores and student perceptions canvassed through questionnaires if it was possible to link individual students' proficiency test scores to their response given in the questionnaire. However, this was not possible due to ethical considerations enforcing protection of anonymity of the participants.

5.7. Recommendations for Further Research

Further research may address a limitation of this study and inaugurate a new line of research by relating learners' test scores to their perceptions. Exploring the link between individual learners' test scores to their perceptions may bring about insights in to washback research. Also, use of longitudinal research designs can be utilized in washback studies since they may better elicit the interaction between instruction, learning and the effects of a test under scrutiny.

In addition, further research may attempt to trace further evidence of the consequential validity of a test by investigating oral (e.g. oral presentations, talks) and written (e.g. reports, essays, projects) real life performances of students.

REFERENCES

Alderson, C. (1998). Testing and teaching: The dream and reality. *Novelty*, 5(4), 23-37.

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

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

Allen, J.P.B., Fröhlich, M. & Spade, N. (1984). The communicative orientation of second language teaching. In J. Handscombe, R. Orem & B. Taylor (Eds.), On TESOL '83 (231-252). Washington, DC: TESOL.

Alderson, C. (2004). Foreword, in Cheng, L., Watanabe, Y. and Curtis, A (Eds). *Washback in language testing: Research Contexts and methods* (pp. 3-18). Manwah, New Jersey: Lawrence Erlbaum Associates.

Andrews, S, Fullilove, J. & Wong, Y. (2002). Targeting washback-a case study. *System,* 20, 207-223.

Angrosino, M. (2007). *Doing Ethnographic and Observational Research*, Sage Publications Ltd., London, UK.

Amrein AL and Berliner DC (2002). High-stakes testing, uncertainty, and student learning. Retrieved 10 February 2010 from http://epaa.asu.edu/epaa/v10n18/

Akpinar, K.D. & Cakildere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and UDS) on productive and receptive skills of academic personnel. *Journal of Language and Linguistic Studies,* 9 (2), 81-94.

Azadi, G. & Gholami, R. (2013). Feedback on washback of EFL tests on ELT in L2 classroom. Theory and Practice in Language Studies, 3(8), 1335-1341.

Bachman, L. (1990). *Fundamental Considerations in Language Testing*, Oxford: Oxford Univesity Press.

Bachman, L. & Palmer, A. (2010). Language Assessment in Practice: Developing language assessments and justifying their use in the real world. Oxford: Oxford. University Press.

Bachman, L. F. & Palmer, A. S. (1996). Language testing in practice. Oxford: Oxford University Press.

Bailey, K. (1999). *TOEFL MonoGraph Series Washback in Language Testing*. Princeton, New Jersey: Educational Testing Service.

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

Baker, E. (1991). *Alternative Assessment and National Policy*. Paper presented at the Natioanl Research Symposium on Limited English Profficent Students' Issues: Focus on Evaluation and Measurement. Washington, DC.

Barbour, R. (2007). *Doing Focus Groups*. London, UK: Sage Publications Ltd.

Bedford, J. (2004). Washback-the effect of assessment on ESOL teaching and learning. ESOL Online Teachers, retrieved from: http://englishonline.tki.org.nz/Media/Files/New-Zealand-Teacher-Research-Studies/Washback-the-Effect-of-Assessment-on-ESOL-Teaching-and-Learning-PDF .

Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods*. Ally & Bacon: Needham Heights, MA.

Brown, H. D. (2003). *Language assessment: Principles and classroom practices.* White Plains, NY: Longman.

Brown, J. D. (2005). *Testing in language program.* New York: McGraw-Hill.

Bryman, A. (Eds.). (2007). *Sage Benchmarks in Social Research Methods Qualitative Research 2 Volume 1 Collecting Data for Qualitative Research*. London, UK: Sage Publications Ltd.

Bryman, A. (Eds.). (2007). *Sage benchmarks in Social Research methods Qualitative Research 2 Volume 2 Quality Issues in Qualitative Research*. London, UK: Sage Publications Ltd.

Bryman, A. (Eds.). (2007). *Sage Benchmarks in Social Research Methods Qualitative Research 2 Volume 3 Issues of Representation and Reflexivity*. London, UK: Sage Publications Ltd.

Bryman, A. (Eds.). (2007). Sage Benchmarks in Social Research Methods Qualitative Research 2 Volume 4 Qualitative Data *Analysis*. London, UK: Sage Publications Ltd.

Burrows, C. (2004). Wash-back in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng, Y. Watanabe, & A. Curtis (Ed.), *Washback in language testing: Research Contexts and methods* (pp. 3-18). Manwah, New Jersey: Lawrence Erlbaum Associates.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and learning. *Applied Linguistics*, 1(1), 1-47.

Chapelle, C., Jamieson, J & Hegelheimer, V. (2003). Validation of a web-based test. *Language Testing*, 20(4), 409-439.

Chapman, D.W. & Synder, C.W. (2000). Can high stakes testing improve instruction: Re-examining conventional wisdom, *International Journal of Educational Development*, 20(6), 457-474.

Chen, L (2002). Washback of a public exam on English teaching. (ERIC Document Reproduction Service No. ED472167).

Cheng, L. (2005). *Changing language teaching through language testing: A washback study.* Cambridge: Cambridge University Press.

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 (Ed.), *Washback in language testing: Research Contexts and methods* (pp. 3-18). Manwah, New Jersey: Lawrence Erlbaum Associates.

Cheng, L. (2004). The washback effect of a public examination change of teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Ed.), *Washback in language testing: Research Contexts and methods* (pp. 147-170). Manwah, New Jersey: Lawrence Erlbaum Associates.

Cheng, L. (2002). The washback effect on classroom teaching of changes in public examinations. In S. J. Savignon (Ed.). Interpreting communicative language teaching: Contexts and concerns in teacher education (pp. 91-111). New York: Yale University Press.

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

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. Language and Education, 11(1), 38-54.

Clapham, C. (2000). Assessment for Academic Purposes: Where next?. *System*, 28, 511-521.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching and assessment.* Cambridge University Press.

Cohen, A. D. (1998). Strategies in learning and using a second language. England: Addison Wesley.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and Perils, *Language Assessment Quarterly*, 10(1), 1-8.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2006). Analysis of discourse features and verification of scoring levels for independent and integrated tasks for the new TOEFL. (TOEFL Monograph No. MS-30 Rm 05–13) Princeton, NJ: ETS

Davies, A. (2003). Three heresies of language testing research. Language Testing, 20(4), 355-368.

Davies, A. (1997). Demands of being professional in language testing. *Language Testing* 14, 328–39

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of Language Testing*. Cambridge University Press

Denzin, N.K. & Lincoln, Y.S. (2005). *The Sage Handbook of Qualitative Research Third Edition*. Sage Publications, Thousand Oaks California, USA.

Djuric, M. (2008) . Dealing with situations of positive and negative washback. *Scprita Manent*, 4 (1), 14-27

D'Agostino, J.V, Welsh, M.E. & Corson, N.M. (2007). Instructional sensitivity of a state's standards based assessment. *Educational Assessment,* 12 (1), p.1-22.

Doe, C., Cheng, L., Fox, J., Klinger, D. & Zheng, Y. (2011). What has experience got to do with it? An exploration of L1 and L2 test takers' perceptions of test performance and alignment to classroom literacy activities. *Canadian Journal of Education*, 34 (3), p.68-85.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review*, 58, 599-622.

Elton, L. R. B., & Laurillard, D. M. (1979). Trends in research on student learning. *Studies in Higher Education*, 4, 87-102.

Farhady, H & Sabeti Dianeti, B. (2000). The Construction and validation of a multiple-choice topic-based English language proficiency. *Roshd Foreign Language Teaching Journal*, 58, p.14-21.

Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Ed.), *Washback in language testing: Research Contexts and methods* (pp. 3-18). Manwah, New Jersey: Lawrence Erlbaum Associates.

Fielding JL, Fielding NG. (2008) 'Synergy and synthesis: integrating qualitative and quantitative data'. in Alasuutari P, Brannen J, Bickman L (eds.) The SAGE Handbook of Social Research Methods London: Sage Article number 33 , 555-571.

Flick, U. (2007). *Designing Qualitative Research*, Sage Publications Ltd., London, UK.

Fredericksen, J. & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27–32.

Frost, K, Elder, C & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test takers' oral performances. *Language Testing*, 20(3), 345-369.

Fox, J. D. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 26-42.

Fox, J., Cheng, L., Berman, R., Song, X. & Myles, Y. (2006). Costs and Benefits: English for Academic Purposes Instruction in Canadian Universities http://www1.carleton.ca/slals/ccms/wp-content/ccms-files/cpals-2006-Fox-et-al.pdf

Fox, J. (2004). Test Decisions over time: Tracking validity. *Language Testing*, 21(4), 437-465.

Fulcher, G. (2000). The `communicative' legacy in language testing. *System*, 28, 483-497.

Gebril, A. & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency, Language assessment Quarterly, 10(1), 9-27.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519–521.

Gibbs, G. (2007). *Analysing Qualitative Data*. Sage Publications Ltd., London, UK

Golombek, P. R. (1998). A Study of Language Teachers' Personal Practical Knowledge. *TESOL Quarterly*, 447-464.

Gosa, C. M. C. (2004). Investigating Washback: A case study using student diaries. Unpublished PhD thesis, Department of Linguistics and Modern English language, Lancaster University, Lancaster, England.

Green, A. (2012) Language Functions Revisited: Theoretical and empirical bases for language construct definition across the ability range. Cambridge: Cambridge University Press. Green, A. (2007a). *IELTS Washback in Context Preparation for Academic Writing in Context.* Cambridge University Press, Cambridge, UK.

Green, A. (2007b). Washback to learning outcomes: a comparative study of IELTS preparation and university professional language courses. *Assessment in Education*, 14(1), 75-97.

Gu, Y. P. (2014). The unbearable lightness of the curriculum: what drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school? *Assessment in Education: Principles, Policy and Practice,* 21(3), p. 286-305.

Guba, E.G. & Lincoln, Y.S. (1989). Fourth generation evaluation. California: Sage Publications.

Haladyna, T., Gale, R. (1981), The role of instructional sensitivity in empirical review of criterion-referenced test items. *Journal of Educational Measurement*, 18 (1), p.39-53.

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

Hamp-Lyons, L. (1998). Ethical test preparation practice: the case of the TOEFL, *TESOL Quarterly*, 33(2), 295-303.

Hanson, R.A., McMorris, R.F., Bailey, J.D (1986). Differences in instructional sensitivity between item formats and between achievement test items. *Journal of Educational Measurement*, 23(1), p. 1-12.

Hayes, B., & Read, J. (2003). IELTS test preparation in New Zealand: preparing students for the IELTS academic module. In R. Tolloh (Ed.), *IELTS Research Report 4* (p. 153-2006). Canberra: IEALTS Australia Pty Limited.

Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: preparing students for the IELTS academic module In L. Cheng, Y. Watanabe, & A. Curtis (Ed.), *Washback in language testing: Research Contexts and methods* (pp. 129-148). Manwah, New Jersey: Lawrence Erlbaum Associates.

Hawkey, R. (2006). *Impact Theory and Practice Studies of the IELTS test and Progetto Lingue 2000.* Cambridge University Press, Cambridge, UK.

Huang, L. (2009). Washback on teacher beliefs and behaviours: Investigating the process from a social psychology perspective. Unpublished PhD thesis, Lancaster University, Lancaster, UK.

Hughes, A. (2003). Testing for language teachers (2nd ed.) Cambridge, CUP.

Hughes, A. (1994). *Backwash and TOEFL 2000.* Unpublished manuscript, commissioned by Educational Testing Service (ETS). University of Reading.

Hughes, A. (1989). *Testing for Language Teachers*. Cambridge, CUP.

Hughes, A. (1988). Introducing a needs-based test of English into an English medium university in Turkey, in Hughes, A. (Eds) *Testing English for University*, Oxford: Modern English Publications, 134-153.

James, M. (2000). Measured lives: the rise of assessment as the engine of change in English schools, *The Curriculum Journal*, 11, 343-364.

Jennings, M., Fox, J., Graves, B., & Shohamy, E. (1999). The test-takers' choice: An investigation of the effect of topic on language-test performance. *Language Testing*, 16(4), 426-456.

Jerry Wellington, M. S. (2007). *Research Methods for the Social Sciences*. London: Continuum International Publishing.

Jordan, R. R. (1997). *English for academic purposes*. Cambridge University Press.

Klein SP, Hamilton LS, McCaffrey DF and Stecher BM (2000). What do test scores in Texas tell us? Education Policy Analysis Archives, 8(49), 4-22.

Krekeler, C. (2006). Language for special academic purposes (LSAP) testing: The effect of background knowledge revisited. *Language Testing*, 23(1), 99-130.

Kosecoff, J. B., & Klein, S. P. (1974, April). Instructional sensitivity statistics appropriate for objectives-based test items. Paper presented at the Annual Conference of the National Council on Measurement in Education, Chicago, IL.

Kvale, S. (2007). *Doing Interviews*. London, UK: Sage Publications Ltd.

Lam, H.P. (1994). Methodology washback and insider's view. *Bringing About Change in Language Education*, 83-99.

Lam, H.P. (1993). Washback-Can it be quantified? A study on the impact of English Examinations in Hong Kong, Unpublished MA dissertation. University of Leeds, UK.

Lightbown, P. M., & Spada, N. (2006). *How Languages are Learned*. New York: Oxford University Press.

MacNamarra, T. (1996). Measuring Second Language Performance, Harlow: Longman.

MacNamarra, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3(1), 31-51.

Madaus, G. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), Critical issues in curriculum: Eighty-seventh yearbook of the National Society for the Study of Education (pp. 83–121). Chicago: University of Chicago Press.

Manjarres, N.B. (2009). Washback of the foreign language test of the state examinations in Colombia: A case study. Arizona working papers in SLAT. Vol, 12, 1-12.

Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologist*, 17, 67-91.

Messick, S. (1982). Validity, in Linn, R.(eds) *Educational Measurement*, NewYork, ACE MacMillian, 13-103.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*. 13(3), 241-256.

Marshall, C & Rossman, G. B. (2011). *Designing Qualitative Research*. California: Sage Publications Inc.

Morrison, K. & Joan, T.F.H. (2002). Testing to destruction: A problem in a small state. *Assessment in Education: Principles, Policy and Practice*, 9(3), 289-317.

Morrow, S.L. (2005). Quality and Trustworthiness in Qualitative Research in Counselling Psychology, *Journal of Counselling Psychology*, 52(2), 250-260.

Morrow, K. (1986). The evaluation of tests of communicative performance in Portal, M. (Eds) *Innovations in Language Testing*, Windsor: NFER-Nelson, 1-13.

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

Muthen, B.O., Kao,C.F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement*, 28(1), 1–22.

Özmen, K.S. (2011). Washback effects of the inter-university foreign language examination on foreign language competencies of candidate academics. Novitas Royal (Research on Youth and Languages), 5(2), 215-228.

Pan, Y. (2011). Teacher washback from English certification exit requirements in Taiwan. Asia Journal of English Language Teaching, 21, 23-42.

Pan, Y. & Newfields, T. (2012). Tertiary EFL proficiency graduation requirements in Taiwan: A study of washback on learning. *Electronic Journal of Foreign Language Teaching*, 9(1), 108-122.

Pertti Alasuutari, L. B. (2008). *The SAGE Handbook of Social Research Methods*. London: SAGE Publishing Inc.

Plakans, L. & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing,* 17, 18-34.

Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. Language Testing, 26(4), 561-587.

Polikoff, M. (2010). Instructional Sensitivity as a Psychometric Property of Assessments. *Educational Measurement*: Issues and Practice, 29 (4), 3–14.

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

Popham, W.J. (1993). Measurement-driven instruction as a 'quick fix' reform strategy, Measurement and Evaluation in Counselling and Development, 26, 31-34.

Popham, W. J. (1978). *Criterion-referenced Measurement*. Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. Journal of Educational Measurement, 6, 1-9.

Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: research contexts and methods* (p.147-170). Mahwah, NJ: Lawrence Erlbaum.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing,* 22(2), 142-173.

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*, 14(1), 51-74.

Rapley, T. (2007). *Doing Conversation, Discourse and Document Analysis*. London: SAGE Publications Ltd.

Rea_Dickins, P. & Scott, C. (2007). Washback from language tests on teaching, learning and policy: Evidence from diverse settings. Assessment in Education: Principles, Policy and Practice, 14(1), 1-7.

Ren, Y. (2011). A study of the washback effects of the College English Test (band 4) on teaching and learning English at tertiary level in China. International Journal of Pedagogies and Learning, 6 (3), 243-259.

Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing,* 23(1), 1-34.

Sawaki, Y., Quinlan, T. & Lee, Y. (2013). Understanding learner strengths and weaknesses: Assessing performance on an integrated writing task. *Language Assessment Quarterly*, 10(1), 73-95.

Shohamy, E. (1992). Beyond proficiency testing: a diagnostic feedback testing model for assessing foreign language learning, *The Modern Language Journal*, 76(4), 513-521.

Shohamy, E. (1993). *The Power of Tests*, Washington DC: National Foreign Language Center.

Shohamy, E. (1998). Critical language testing. *Studies in Educational Evaluation*, 24(4), 331-345.

Shohamy, E. (2001). *The Power of Tests*: *A critical perspective of the uses of language tests*, Harlow: Longman.

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

Smith, M.L. (1991). Meanings of test preparation, American Educational Research Journal, 28(3), 521-542.

Smithson, J. (2008). *Focus Groups The Sage Handbook of Social Research Methods*, (Ed.) Alasuutari, P., Bickman, L. & Brannen, J., p. 358-370, London, UK: Sage Publications Ltd.

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

Scott, C. (2007). Stakeholder perceptions of test impact. *Assessment in Education*, 14 (1), 27-49.

Shih, C. (2014). The washback of the General English proficiency Test on university policies: A Taiwanese case study. *Language Assessment Quarterly*, 7 (3), 234-254.

Spada, N. & M. Fröhlich. 1995. Communicative orientation of language teaching observation scheme. Coding conventions and applications. Sydney: NCELTR, Macquarie University.

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.

Taylor, L. (2005). Washback and Impact. ELT Journal, 59(2), 154-155.

Tratnik, A. (2008). Key issues in testing English for specific purposes. *Scripta Manent*, 4(1), 3-13.

Turner, C.E. (2009). Examining washback in second language education contexts: A high-stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *International Journal of Pedagogies and Learning*, 5, 103-123.

Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices, selected papers from the 19th ISTAL.

Tsagari, D. (2009). The complexity of test washback. Frankfurt am Main: Peter Lang.

Tsagari, D. (2007) Review of Washback in Language Testing: What Has Been Done? What More Needs Doing? Washington, DC: Center for Applied Linguistics (ERIC Document Reproduction Services No. ED 497709). http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED497709&ERICExtSearch_SearchType_0=eric_accno&accno=ED497709

Tsagari, D. (2006). Investigating the washback effect of a high-stakes EFL exam in the Greek context: Participants' Perceptions, materials design and classroom applications. Unpublished PhD thesis, Lancaster University, Lancaster, UK.

Turner, C.E. (2009). Examining washback in second language education contexts: A high stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *The International Journal of Pedagogies and Learning*, 5, 103-123.

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

Wall, D. & Horak, T. (2007). Using baseline studies in the investigation of test impact. Assessment in Education: Principles, Policy and Practice, 14 (1), 99-116.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge: Cambridge University Press.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28, 499-509.

Watanabe, Y. (2004a). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Ed.), *Washback in language testing: Research Contexts and methods* (pp. 19-36). Manwah, New Jersey: Lawrence Erlbaum Associates.

Watanabe, Y. (2004b). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis (Ed.), *Washback in language testing: Research Contexts and methods* (pp. 129-148). Manwah, New Jersey: Lawrence Erlbaum Associates.

Watanabe, Y. (2000). Washback effects of the English section of the Japanese university entrance examinations on instruction in-pre-college level EFL. Language Testing Update, 27, p. 42-47.

Watanabe, Y. (1997). *The washback effects of the Japanese university entrance examinations of English-classroom-based research*, unpublished PhD thesis, University of Lancaster, Lancaster.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from class-rom based research, *Language Testing*, 13(3), 318-333.

Watanabe, Y. (1992). Washback effects of College Entrance Examination on language learning strategies. JACET, 175-194.

Weigle, S., Yang, W. & Montee, M. (2013). Exploring reading processes in an academic reading test using short answer questions. *Language Assessment Quarterly*, 10(1), 28-48.

Weigle, S. & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21, 118-133.

Weili, W. (2010). Investigating the washback effect of the new CET 4 listenign comprehension subtest on language learners. Chines eJournal of Applied Linguistics, 33(5), 28-38.

Weir, C. (1990). *Communicative Language Testing,* Hemel Hempstead: Prentice Hall.

Wolfersberger, M. (2013). Refining Construct of classroom-based writing-from-readings assessment: The role of task representation. Language Assessment Quarterly, 10(1), 49-72.

Xiao, W. (2014). The intensity and direction of CET Washback on Chinese college students' test-taking strategy use. *Theory and Practice in Language Studies*, 4(6), 1171-1177.

Xie, Q. (2013). Does test prperartion work? Implications for score validity. *Language Assessment Quarterly*, 10 (2), 196-218.

Xie, Q. & Andrews, S (2012). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modelling. *Language Testing*, 1-22.

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. *Asian EFL Journal*, 12(2), p. 92-116.

Yu, G. (2013). From integrative to integrated language assessment: Are we there yet?. *Language Assessment Quarterly,* 10(1), 100-114.

Zhan, Y. & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self-theories. Assessment in Education, 21(1), 71-89

APPENDICES

STUDY OF THE WASHBACK AND INSTRUCTIONAL SENSITIVITY OF THE TRACE
EXAM STUDENT QUESTIONNAIRE

Dear Students,
As a part of my PhD research I need your help in completing this questionnaire.

Purpose of the questionnaire
The main purpose of this questionnaire is to investigate students' perceptions in (1) how TRACE affects their learning and (2) how much they think they learned upon the completion of the preparatory program.

Confidentiality
The information you provide in the questionnaire will remain completely confidential. You may respond anonymously if you wish.If you have any questions or concerns please e-mail Asli Saglam : asli.saglam@ozyegin.edu.tr. Thanks in advance

Thank you for agreeing to complete the questionnaire. To confirm your agreement, please sign below.
Signature: …………………………   Date:………………………….

Part 1 ENGLISH LANGUAGE BACKGROUND: For questions 1-2, Put a tick (√) under the relevant box to answer and for questions 3-5 write down in boxes.

	Yes	No
1. As a child, did you live with a parent, guardian or other close relation who was a native speaker of English?		

2. Did you study English in:	Yes	No
☐ Kindergarten (age 3-6)?		
☐ Primary school (age 7-11)?		
☐ Secondary school (age 12-17)?		
☐ University?		
☐ Extra language classes?		

3. Have you studied in English speaking countries?		
If **yes**, where _____, when_____ and for how long_____?		

4. At which level have you started the prep program?	

PART 2 ATTITUDES TO TEACHING, MATERIALS, TASKS AND TRACE

A. Please put a tick (√) in the column that best describes your ideas about TRACE	Strongly Disagree	Disagree	Agree	Strongly Agree
10. I understand what English language and academic skills are required by TRACE.				
11. I think the skills tested in TRACE are necessary to learn				
12. I think TRACE tests what I will be asked to do at my department after prep school.				
13. I feel TRACE is an appropriate test.				
14. I think TRACE measures my language ability level and academic skills accurately.				
15. I think students should put in extra effort to prepare themselves for TRACE outside class.				
16. I think TRACE has affected the way my teacher designs our lessons.				
B. Please put a tick (√) in the column which shows how often you do the following tasks in your English classes.	Never	Sometimes	Frequently	Always
20. discussions with a partner				
21. discussions in small groups				
22. writing essays				
23. self and/or peer evaluation of essays using criteria given by the teacher				
24. games, puzzles, quizzes				
25. group interviews, and other project work				
26. creative writing				
27. web/on-line activities				
28. role plays				
29. grammar exercises				
30. skills training (how to read, listen, speak, learn vocabulary and write better)				
31. vocabulary exercises				
32. exam practice				
33. listening to lectures and taking notes to answer questions				
34. reading a text and answering questions				
35. reading texts from books and other materials to use information in writing an essay				
36. reading different texts and answering questions that compare these different reading texts				
37. using notes (from reading or listening) to respond to a writing question in essay form				
38. working on skills and strategies about how to take a test				
If you had any other activities that are not listed above, please list them here				

C. Please put a tick (√) in the column which shows what you think of the materials (books and supplementary materials) used in class	Strongly Disagree	Disagree	Agree	Strongly Agree
1. Our materials covered **reading skills** such as the following that helped me to learn English well:				
1.1 understanding main ideas				
1.2 understanding the writer's overall point,				
1.3 inferring purpose, tone, and attitude of the writer				
1.4 guessing meanings of unknown words				
1.5 analysing organisation of the text				
1.6 synthesizing information from different texts to answer questions				
2. Our materials covered reading skills that would help me to perform well on the test.				
3. There was a variety of reading text types (short texts and long texts)				
4. Reading texts were followed by **different activities** such as:				
4.1 multiple choice,				
4.2 summarising				
4.3 matching				
4.4 open-ended questions				
4.5 fill-in blanks				
4.6 choose the sentence that completes a paragraph				
4.7 write or choose a title for the text				

If you have any other comment on question techniques, text types and reading skills covered or not covered by the materials you had in class, please write them here:

	Strongly Disagree	Disagree	Agree	Strongly Agree
5. Our materials covered **listening skills** such as the following that helped me learn English well				
5.1 understanding overall point				
5.2 inferring purpose, tone, and attitude of the speaker(s)				
5.3 guessing meanings of unknown words				
5.4 noting down information and using these notes to answer questions				
5.5 analysing the organisation of the listening text to understand main and supporting ideas				
5.6 combining my listening notes with other sources to write an essay.				
6. Our materials covered listening skills that would help me perform well on the test.				

7. There was a variety of listening text types (e.g. conversations, talks, TV shows and lectures)				
8. Listening texts were followed by different activities such as:				
8.1 multiple choice,				
8.2 summarising				
8.3 matching				
8.4 open-ended questions				
8.5 fill-in the blanks				

If you have any other comment on question techniques, text types and listening skills covered or not covered by the materials you had in class, please write them here:

	Strongly Disagree	Disagree	Agree	Strongly Agree
9. Our materials covered **writing skills** such as the following that helped me learn English well				
9.1 organising ideas from my notes into a written plan				
9.2 introducing and narrowing down ideas				
9.3 organising my ideas in a logical way				
9.4 using information and arguments from different sources into my writing				
9.5 evaluating my own work and revising				
9.6 evaluating my peers' writing based on criteria given by the teacher				
10. Our materials covered writing skills that would help me perform well on TRACE.				
11. There was a variety of opportunities for writing including:				
11.1 writing reviews				
11.2 creative writing				
11.3 report writing				
11.4 writing a response to a text				
11.5 story writing				
11.6 summary writing				
12. I made use of information given in reading and listening texts in the writing activities.				

If you have any other comment on question techniques, text types and writing skills covered or not covered by the materials you had in class, please write them here:

	Strongly Disagree	Disagree	Agree	Strongly Agree
13. Our materials covered speaking skills such as the following that helped me learn English well:				
13.1 making polite requests				
13.2 starting and ending discussions in an appropriate way				
13.3 explaining my ideas				
13.4 giving examples				
13.5 agreeing and disagreeing				
13.6 asking for clarification questions (e.g. Could you explain your words?)				
14. There was a variety of opportunities and activities for speaking including:				
14.1 pair work discussions				
14.2 role-play				
14.3 group work discussions				
14.4 presentation				
14.5 asking and answering questions				
14.6 story telling				
14.7 giving talks				
15. I made use of information given in reading and listening texts in the speaking activities.				
If you have any other comment on activities, text types and speaking skills covered or not covered by the materials you had in class, please write them here:				

	Never	Sometimes	Frequently	Always
16. Our materials covered **grammar & vocabulary** that are important for learning English				
17. Our materials covered **grammar & vocabulary** that are important to do well on tests.				
18. There was a variety of **grammar & vocabulary** activities such as games, writing…etc.				
19. Grammar & vocabulary lessons were followed by different activities such as:				
19.1 multiple choice				
19.2 matching				
19.3 fill-in the blanks				
19.4 using newly learned grammar and vocabulary in writing				
19.5 using newly learned grammar and vocabulary in speaking				
If you have any other comment on question techniques and language skills (grammar & vocabulary) covered or not covered by the materials you had in class, please write them here:				

PART 3. INSTRUCTIONAL SENSITIVITY - HOW MUCH LEARNING HAPPENED?

Please put a tick (√) in the column that best describes your ideas about How much you learned in your class in their level (in the preparatory program)	Strongly disagree	Disagree	Agree	Strongly Agree
Reading: I learned how to…				
25. Find out the main idea(s) of the text.				
26. Identify supporting ideas, details and examples.				
27. Find out the message of the text by looking at visual (layout, headings) and contextual clues (choice of words, grammar structures, examples given) in the text.				
28. Identify specific information by using some strategies (such as reading carefully, taking reading notes, leaving out some information if I didn't understand and tried to guess the meaning, searching for info quickly).				
29. Relate ideas in a reading to one another by understanding how the text is organised, whether ideas follow one another, whether the writer explains his ideas by explanations and examples.				
30. Guess the meaning of unknown words using my knowledge.				
31. Read between the lines to understand the author's attitude, tone and purpose that are not openly stated in the text.				
32. Understand words and phrases such as "**This** shows …, **These** are…" that refer to objects, people or concepts.				
33. Recognize different organisations of ideas (definition, cause and effect, comparison…etc.) in the reading text to understand the author's overall point of view.				
34. Select and make notes from reading texts to answer questions.				
35. Combine my reading notes with other sources to write an essay.				
36. Compare and contrast different texts to understand why they are written (purpose), have different styles (tone, author's attitude), for whom the text is written (audience)				
	Strongly disagree	Disagree	Agree	Strongly Agree
Listening: I learned how to…				
31. Guess the meaning of new words to understand a listening text				
32. Find out the main ideas based on the clues in the listening text such as organisation of the listening text, examples and key words used and questions asked by the speaker(s).				
33. Identify specific information (by paying attention to names, dates, numbers, signal words such as one of the important problems…, firstly,…etc.)				
34. Understand the relationship between ideas				

	Strongly disagree	Disagree	Agree	Strongly Agree
35. Understand the speaker's use of stress and intonation to signal different sections of a listening text.				
36. Recognize organisation of different sections of a listening text.				
37. Select and take notes from an organised listening text.				
38. Use my listening notes to answer questions.				
39. Combine my listening notes with other sources to write an essay.				
40. Infer who speakers are, what the situation is, speaker's attitude and purpose				
41. Follow the arguments of talks and presentations if they are organised and on a familiar topic.				
	Strongly disagree	**Disagree**	**Agree**	**Strongly Agree**
Writing: I learned how to…				
31. Give short answers to the questions about the texts I read or listen to				
32. Complete missing info about the texts I read or listen to				
33. Take organised notes while listening				
34. Take organised notes while reading				
35. Organise my ideas in an essay, generate content, link my ideas to one another in paragraphs and in the whole essay by explaining, using linking words (e.g. In addition, on the other hand, however, furthermore…etc.) and giving examples.				
36. Generate content, organise my ideas in to a plan/an outline, introduce the topic and relate my ideas to each other by using linking words				
37. Support my ideas using explanations, descriptions, statistical data, information from other sources and examples				
38. Conclude my essay by summarising main points and offering my final comments on the topic.				
39. Choose appropriate organisational patterns (such as defining concepts, comparing and contrasting, describing cause and effects …etc.) to develop my ideas in the paragraphs				
40. Write essays of about 400-450 words in which I can organize and express my ideas well and use information from other sources.				
41. Summarise or paraphrase information from other sources such as articles and lectures				
42. Re-write my essay by getting feedback from my class mates				
43. Re-write my essay by getting feedback from my teacher				
44. Edit my own work by self-evaluating it				
45. Write in an academic style by using mostly correct grammar, spelling and punctuation, as well as variety of grammar structures and appropriate vocabulary.				
	Strongly disagree	**Disagree**	**Agree**	**Strongly Agree**

Language: I learned how to…				
1. Record and remember newly learned vocabulary items in an organised way				
2. Use newly learned vocabulary in my speaking				
3. Use newly learned vocabulary in my writing				
4. Use dictionaries effectively				
5. Use different grammar structures accurately				

APPENDIX B

STUDY OF THE WASHBACK AND INSTRUCTIONAL SENSITIVITY OF THE TRACE EXAM

TEACHER QUESTIONNAIRE

Dear colleagues, As a part of my PhD research I need your help in completing this questionnaire. **Purpose of the questionnaire** The main purpose of this questionnaire is to investigate teachers' perceptions in (1) how TRACE affects their teaching methodology and (2) how much they think their students learned upon the completion of the preparatory program. **Confidentiality** The information you provide in the questionnaire will remain completely confidential. You may respond anonymously if you wish. If you have any questions or concerns please e-mail Asli Saglam : asli.saglam@ozyegin.edu.tr Thanks in advance Asli Lidice Gokturk Saglam, Instructor	Thank you for agreeing to complete the questionnaire. To confirm your agreement, please sign below. Signature: …………………………… Date:…………………………….

PART 1 ATTITUDES TO TEACHING, MATERIALS, TASKS AND TRACE

A. Please put a tick (√) in the column that best describes your ideas about TRACE	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I understand what English language and academic skills are required by TRACE.				
2. I think the skills tested in TRACE are necessary to teach				
3. I think TRACE tests what students will be asked to do at their departments after prep school.				
4. I feel that TRACE is an appropriate test.				
5. I think TRACE measures my students' language ability level and academic skills accurately.				

	Never	Sometimes	Frequently	Always
6. I think students should put in extra effort to prepare themselves for TRACE outside class.				
7. I think TRACE has affected the way I design my lessons.				
B. Please put a tick (√) in the column which shows how often you do the following tasks in your English classes	**Never**	**Sometimes**	**Frequently**	**Always**
1. discussions with a partner				
2. discussions in small groups				
3. writing essays				
4. self and/or peer evaluation of essays using criteria given by the teacher				
5. games, puzzles, quizzes				
6. group interviews, and other project work				
7. creative writing				
8. web/on-line activities				
9. role plays				
10. grammar exercises				
11. skills training (how to read, listen, speak, learn vocabulary and write better)				
12. vocabulary exercises				
13. exam practice				
14. listening to lectures and taking notes to answer questions				
15. reading a text and answering questions				
16. reading texts from books and other materials to use information in writing an essay				
17. reading different texts and answering questions that compare these different reading texts				
18. using notes (from reading or listening) to respond to a writing question in essay form				
19. working on skills and strategies about how to take a test				

If you had any other activities that are not listed above, please list them here

	Strongly Disagree	Disagree	Agree	Strongly Agree
C. Please put a tick (√) in the column which shows what you think of the materials (books and supplementary materials) used in class	**Strongly Disagree**	**Disagree**	**Agree**	**Strongly Agree**
1. Our materials covered **reading skills** such as the following that helped my students to learn English well:				
1.1 understanding the main ideas				
1.2 understanding the writer's overall point,				

1.3 inferring purpose, tone, and attitude of the writer			
1.4 guessing meanings of unknown words			
1.5 analysing organisation of the text			
1.6 synthesizing information from different texts to answer questions			
2 Our materials covered reading skills that would help my students to perform well on the test.			
3 There was a variety of reading text types (short texts and long texts)			
4 Reading texts were followed by **different activities** such as:			
4.1 multiple choice,			
4.2 summarising			
4.3 matching			
4.4 open-ended questions			
4.5 fill-in blanks			
4.6 choose the sentence that completes a paragraph			
4.7 write or choose a title for the text			

If you have any other comment on question techniques, text types and reading skills covered or not covered by the materials you had in class, please write them here:

	Strongly Disagree	Disagree	Agree	Strongly Agree
5 Our materials covered **listening skills** such as the following that helped my students learn English well:				
5.1 understanding overall point				
5.2 inferring purpose, tone, and attitude of the speaker(s)				
5.3 guessing meanings of unknown words				
5.4 noting down information and using these notes to answer questions				
5.5 analysing the organisation of the listening text to understand main and supporting ideas				
5.6 combining my listening notes with other sources to write an essay.				
6 Our materials covered listening skills that would help my students perform well on the test.				
7 There was a variety of listening text types (e.g. conversations, talks, TV shows and lectures.				
8 Listening texts were followed by different activities such as:				
8.1 multiple choice,				

8.2 summarising				
8.3 matching				
8.4 open-ended questions				
8.5 fill-in the blanks				

If you have any other comment on question techniques, text types and listening skills covered or not covered by the materials you had in class, please write them here:

	Strongly Disagree	Disagree	Agree	Strongly Agree
9 Our materials covered **writing skills** such as the following that helped my students learn English well.				
9.1 organising ideas from my notes into a written plan				
9.2 introducing and narrowing down ideas				
9.3 organising my ideas in a logical way				
9.4 using information and arguments from different sources into my writing				
9.5 evaluating my own work and revising				
9.6 evaluating my peers' writing based on criteria given by the teacher				
10. Our materials covered writing skills that would help my students perform well on TRACE.				
11. There was a variety of opportunities for writing including:				
11.1 writing reviews				
11.2 creative writing				
11.3 report writing				
11.4 writing a response to a text				
11.5 story writing				
11.6 summary writing				
12. My students made use of information given in reading and listening texts in the writing activities.				

If you have any other comment on question techniques, text types and writing skills covered or not covered by the materials you had in class, please write them here:

	Strongly Disagree	Disagree	Agree	Strongly Agree
13. Our materials covered speaking skills such as the following that helped my students to learn English well:				
13.1 making polite requests				
13.2 starting and ending discussions in an appropriate way				
13.3 explaining my ideas				
13.4 giving examples				
13.5 agreeing and disagreeing				
13.6 asking for clarification questions (e.g. Could you explain your words?)				
14. There was a variety of opportunities and activities for speaking including:				
14.1 pair work discussions				
14.2 role-play				
14.3 group work discussions				
14.4 presentation				
14.5 asking and answering questions				
14.7 story telling				
13.7 giving talks				
15. My students made use of information given in reading and listening texts in the speaking activities.				

If you have any other comments on activities, text types and speaking skills covered or not covered by the materials you had in class, please write them here:

	Never	Sometimes	Frequently	Always
46. Our materials covered **grammar & vocabulary** that are important for learning English				
47. Our materials covered **grammar & vocabulary** that are important to do well on tests.				
48. There was a variety of **grammar & vocabulary** activities				
49. Grammar & vocabulary lessons were followed by different activities such as:				
19.1 multiple choice				

	Strongly disagree	Disagree	Agree	Strongly Agree
19.2 matching				
19.3 fill-in the blanks				
19.4 using newly learned grammar and vocabulary in writing				
19.5 using newly learned grammar and vocabulary in speaking				
If you have any other comment on question techniques and language skills (grammar & vocabulary) covered or not covered by the materials you had in class, please write them here:				

PART 3. INSTRUCTIONAL SENSITIVITY - HOW MUCH LEARNING HAPPENED?

Please put a tick (√) in the column that best describes your ideas about How much your students learned in your class in their level (in the preparatory program)	Strongly disagree	Disagree	Agree	Strongly Agree
Reading: My students learned how to…				
1. Find out the main idea(s) of the text.				
2. Identify supporting ideas, details and examples.				
3. Find out the message of the text by looking at visual (layout, headings) and contextual clues (choice of words, grammar structures, examples given) in the text.				
4. Identify specific information by using some strategies (such as reading carefully, taking reading notes, leaving out some information if they don't understand and tried to guess the meaning, searching for info quickly).				
5. Relate ideas in a reading to one another by analysing how the text is organised, whether ideas follow one another, whether the writer explains his ideas by explanations and examples.				
6. Guess the meaning of unknown words using my knowledge.				
7. Read between the lines to understand author's attitude, tone and purpose that are not openly stated in the text.				
8. Understand words and phrases such as "**This** shows …, **These** are…" that refer to objects, people or concepts.				
9. Recognize different organisations of ideas (definition, cause and effect, comparison…etc.) in the reading text to understand the author's overall point of view.				
10. Select and make notes from reading texts to answer questions.				
11. Combine their reading notes with other sources to write an essay.				

	Strongly disagree	Disagree	Agree	Strongly Agree
12. Compare and contrast different texts to understand why they are written (purpose), have different styles (tone, author's attitude), for whom the text is written (audience)				

Listening: My students learned how to…				
1. Guess the meanings of new words to understand a listening text				
2. Find out the main ideas based on the clues in the reading such as organisation of the listening text, examples and key words used and questions asked by the speaker(s).				
3. Identify specific information (by paying attention to names, dates, numbers, signal words such as (one of the important problems…, firstly,…etc.)				
4. Understand the relationship between ideas.				
5. Understand the speaker's use of stress and intonation to signal different sections of a listening text.				
6. Recognize organisation of different sections of a listening text.				
7. Select and take notes from an organised listening text.				
8. Use their listening notes to answer questions.				
9. Combine their listening notes with other sources to write an essay.				
10. Infer who speakers are, what the situation is, speaker's attitude and purpose.				
11. Follow the arguments of talks and presentations if they are organised and on a familiar topic.				

	Strongly disagree	Disagree	Agree	Strongly Agree
Writing: My students learned how to…				
1. Give short answers to the questions about the texts they read or listen to.				
2. Complete missing info about the texts they read or listen to.				
3. Take organised notes while listening.				
4. Take organised notes while reading.				
5. Organise their ideas in an essay, generate content, link my ideas to one another in paragraphs and in the whole essay by explaining, using linking words (e.g. In addition, on the other hand, however, furthermore…etc.) and giving examples.				
6. Generate content, organise my ideas in to a plan/an outline, introduce the topic and relate their ideas to each other by using linking words.				

	Strongly disagree	Disagree	Agree	Strongly Agree
7. Support their ideas using explanations, descriptions, statistical data, information from other sources and examples.				
8. Conclude their essay by summarising main points and offering my final comments on the topic.				
9. Choose appropriate organisational patterns (such as defining concepts, comparing and contrasting, describing cause and effects …etc.) to develop their ideas in the paragraphs.				
10. Write essays of about 400-450 words in which they can organize and express their ideas well and use information from other sources.				
11. Summarise or paraphrase information from other sources such as articles and lectures.				
12. Re-write their essay by getting feedback from their class mates.				
13. Re-write their essay by getting feedback from their teacher.				
14. Edit their own work by self-evaluating it.				
15. Write in an academic style by using mostly correct grammar, spelling and punctuation, as well as variety of grammar structures and appropriate vocabulary.				
	Strongly disagree	**Disagree**	**Agree**	**Strongly Agree**
Language: My students learned how to…				
1. Record and remember newly learned vocabulary items in an organised way				
2. Use newly learned vocabulary in their speaking				
3. Use newly learned vocabulary in their writing				
4. Use dictionaries effectively				
5. Use different grammar structures accurately				

Thanks a lot for completing the questionnaire

APPENDIX C

SEMI STURUCTURED STUDENT FOCUS GROUP INTERVIEW TOPICS AND
QUESTIONS FOR STUDY OF THE WASHBACK AND INSTRUCTIONAL SENSITIVITY
OF THE TRACE

0. Opening
Introduction
Key points of the study, purpose, confidentiality, media and timing

1. Instructional sensitivity of TRACE
How much learning takes place in terms of:

Reading

1. Do you think you have improved your reading ability in the course? Why? Why not?
2. What do you think you learned in terms reading skills in the course?   Can you give some examples?

Listening

3. Do you think you have improved your listening ability in the course? Why? Why not?
4. What do you think you have learned in terms of listening skills? Can you give some examples?

Writing

5. Do you think that you improved yourself in writing?
6. What do you think you have learned in terms of writing skills? Can you give some examples?

2. WASHBACK

Attitudes to Materials & Tasks

7. Think about the course materials (books, supplementary materials, web activities..etc.) Do you think that they have contributed to your learning English? Which ones were the most beneficial in your opinion? Why?

Coursebook

Do you think that the difficulty level of course book was suitable for the level?

Do you think that it has contributed to your learning English?

Do you think that the book prepares you for TRACE?

Supplementary materials:

Do you think that they have contributed to your learning English?

Do you think that they prepare you for TRACE?

Vocabulary & Grammar Booklet:

Do you think that the difficulty level of course book was suitable for the level?

Do you think that it has contributed to your learning English?

Do you think that the book prepares you for TRACE?

Blended Learning Web Materials

Do you think that the difficulty level was suitable for the level?

Do you think that they have contributed to your learning English?

Do you think that they prepare you for TRACE?

8. What kind(s) of reading, listening & writing activities and tasks have you done in the class?
9. Do you remember any task that was directly related to the test and it may help you improve your scores?

Correspondence between teaching-learning and being successful on TRACE (The relationship between Objectives of the program, learning and TRACE)

10. Do you think that content of the course (what you learned in class) and TRACE are similar? How?
11. How do you think what you learned in the course may help you in TRACE?
12. In your opinion to what extent did the course support you to learn English and be successful TRACE? How well did the course prepare you to be successful on TRACE?

3. Round up and thanks

APPENDIX D

SEMI-STRUCTURED TEACHER INTERVIEW TOPICS AND QUESTIONS FOR THE
STUDY OF WASHBACK AND ISNTRUCTIONAL SENSITIVITY OF TRACE

1. Opening

   Introduction
   Key points of the study, purpose, confidentiality, media and timing

2. Instructional sensitivity of TRACE

   How much learning takes place in terms of:
   Reading

1. Do you think your students have improved their reading ability in the course? Why? Why
   not?
2. What do you think they have learned in terms of reading skills in the course? Can you
   give some examples?
   Listening
3. Do you think your students have improved their listening ability in the course? Why?
   Why not?
4. What do you think they have learned in terms of listening skills in the course? Can you
   give some examples?
   Writing
5. Do you think that your students have improved themselves in writing?
6. What do you think they have learned in terms of writing skills in the course? Can you
   give some examples?

   WASHBACK: Attitudes to Teaching materials & Tasks

7. Think about the course materials (books, supplementary materials, web activities..etc.)
   Do you think that they have contributed to your students' learning English? Which ones
   were the most beneficial in your opinion? Why?

Coursebook:

   Is the course book a B2 level book?
   Does the book prepare your students for TRACE?

Supplementary materials:

   Are the supplementary materials used for TRACE? Why?
   Would you use a supplementary material if you didn't have to prepare students for
   TRACE?

Vocabulary & Grammar Booklet:

   Do they prepare students for TRACE?
   Do you think that they have contributed to your students' learning English?

Blended Learning Web Materials

   Do they prepare students for TRACE?

Do you think that they have contributed to your students' learning English?

8. What kind(s) of reading, listening & writing activities and tasks have you done in the class?
9. Do you think that your selection of activities and tasks are affected by TRACE?
10. Do you remember any task that was directly related to the test and it may help your students improve their scores?

Correspondence between teaching-learning and being successful on TRACE (The relationship between Objectives of the program, learning and TRACE)

11. Do you think that content of the course (what you teach in class) and TRACE are similar? How?
12. Do you do any special preparation for TRACE? If not then: How do you think what your students learned in the course may help them in TRACE?
13. What changes would you initiate in your teaching if your students didn't take TRACE at the end of the year?
14. In your opinion to what extent did the course support your students to learn English and be successful on TRACE? How well did the course prepare them to be successful on TRACE?

Round up and thanks

APPENDIX E

STUDY OF THE VIABILITY OF TRACE BASED TEST DECISIONS OVER TIME TEACHER QUESTIONNAIRE

Dear colleagues,	Thank you for agreeing to complete the questionnaire.
As a part of my PhD research I need your help in completing this questionnaire. **Purpose of the questionnaire** The main purpose of this questionnaire is to investigate how teachers perceive the viability of TRACE based test decisions over time and how much you think your freshman students learned in terms of language competency and academic skills upon the completion of the preparatory program in English language. **Confidentiality** The information you provide in the questionnaire will remain completely confidential. You may respond anonymously if you wish. If you have any questions or concerns please e-mail Asli Saglam : asli.saglam@ozyegin.edu.tr. Thanks in advance	To confirm your agreement, please sign below. Signature: ………………………… Date:…………………………….

Gender: **Male**   **Female**

Major (graduate and post graduate) (TESOL, Literature, EFL, Applied Linguistics, Education…etc.) _____

Years of work experience   **1-5** ☐   **6-10** ☐   **11-15** ☐   **16-20** ☐   **More than 20 years** ☐

 Highest Qualification Achieved: _____

At the end of the Preparatory School Program in English language, our students take the Test of Readiness of Academic English (TRACE). TRACE results are assumed to be linked to the use of English for Academic Purposes at mainstream departmental university courses. We would appreciate your answers to the following questions. Take as much time as you need.

1) Which course(s) do you teach at UG Department? _____

230

2) Do you think students who completed Preparatory School Program are well-prepared regarding their English language ability and academic skills training? Check the degree in the table below that reflects your point of view.

	not prepared	fairly prepared	prepared	well prepared
1. Reading academic texts and understanding the main ideas				
2. Taking reading notes				
3. Understanding lectures				
4. Taking listening notes				
5. Writing an organised essay				
6. Discussing ideas and expressing opinions clearly and accurately in their speech				
7. Asking questions				
8. Using a range of vocabulary appropriately				
9. Using a range of grammatical structures in their written and spoken work				
10. Using different sources (notes, summaries etc.) to support ideas in their written and spoken work				
11. Giving feedback to peers				
12. Revising own written work based on given feedback				
If you have further comments about any other English language ability and academic skills, please write them here:				

3) **Where do you see the strengths and weaknesses of the students who completed Preparatory School Program? Please comment under relevant headings in the table below.**

	Strengths of the students who completed Preparatory School Program	Weaknesses of the students who completed Preparatory School Program
In Writing		

In Reading		
In Listening		
In Speaking		
In using Grammar and Vocabulary		
If you have any further comments about any other **use of English for Academic Purposes at mainstream departmental university courses please write them here.**		

4) What do you suggest to improve the Preparatory School Program?

Thanks a lot for completing the questionnaire

APPENDIX F

STUDY OF THE VIABILITY OF TRACE BASED TEST DECISIONS OVER TIME QUESTIONNAIRE FOR FRESHMAN STUDENTS

Dear Students, As a part of my PhD research I need your help in completing this questionnaire. **Purpose of the questionnaire** The main purpose of this questionnaire is to investigate what students think about the viability of TRACE based test decisions over time and how much you think you learned in terms of language competency and academic skills upon the completion of the preparatory program in English language **Confidentiality** The information you provide in the questionnaire will remain completely confidential. You may respond anonymously. If you have any questions or concerns please e-mail Asli Saglam : asli.saglam@ozyegin.edu.tr. Thanks in advance	Thank you for agreeing to complete the questionnaire. To confirm your agreement, please sign below. Signature: …………………………… Date:…………………………….

At the end of the Preparatory School Program in English language, our students take the Test of Readiness of Academic English (TRACE). TRACE results are assumed to be linked to the use of English for Academic Purposes at mainstream departmental university courses. We would appreciate your answers to the following questions. Take as much time as you need.

1) What is your department? _____

2) Do you think students who completed Preparatory School Program are well-prepared regarding their English language ability and academic skills training? Check the degree in the table below that reflects your point of view.

	not prepared	fairly prepared	prepared	well prepared
1. Reading academic texts and understanding the main ideas				
2. Taking reading notes				
3. Understanding lectures				
4. Taking listening notes				
5. Writing an organised essay				
6. Discussing ideas and expressing opinions clearly and accurately in their speech				
7. Asking questions				
8. Using a range of vocabulary appropriately				
9. Using a range of grammatical structures in their written and spoken work				
10. Using different sources (notes, summaries etc.) to support ideas in their written and spoken work				
11. Giving feedback to peers				
12. Revising own written work based on given feedback				
If you have further comments about any other English language ability and academic skills, please write them here:				

3) **Where do you see the strengths and weaknesses of the students who completed Preparatory School Program? Please comment under relevant headings in the table below.**

	Strengths of the students who completed Preparatory School Program	Weaknesses of the students who completed Preparatory School Program
In Writing		

In Reading		
In Listening		
In Speaking		
In using Grammar and Vocabulary		
If you have any further comments about any other **use of English for Academic Purposes at mainstream departmental university courses please write them here.**		

4) What do you suggest to improve to Preparatory School Program in order to prepare you better in the **use of English for Academic Purposes later on at mainstream departmental university courses**?

Thanks a lot for completing the questionnaire

One-on-One Interview Topics for Study of the Consequential Validity

0. Introduction

Key points of the study; purpose, confidentiality, media and timing

1. <u>Correspondence between exit criteria of TRACE and expected academic skills at the departments</u>

   Could you share your observations regarding student achievement in:

   In Writing

   In Reading

   In Listening

   In Speaking

   In use of Grammar and Vocabulary

   In academic skills

2. What skills are required at the department?

3. Do you think that students who complete Prep Program can cope with the academic demands of your department

4. What needs to be done in prep program to better equip these students for their departments?

5. <u>Other Views</u>

6. <u>Round up and Thanks</u>

CLASSROOM OBSERVATION SCHEME

CLASSROOM OBSERVATION NOTE-TAKING SHEET																	
Date:									*Number of Students:*								
Teaching objectives:									*Teacher:*								
Course level:									*(Back ground):*								
Time:																	
Time From -	**Time To**	**Activity**		**Materials**				**Skill/ Knowledge focus**							**Student Modality**			**Comments**
			C B	**S**	**SP**	**O**	**L**	**S**	**R**	**W**	**I**	**G**	**V**	**I**	**P**	**G**	**C**	

Materials: CB (Course Book-Commercialised), S (Supplementary Materials), SP (Self prepared by the teacher), O(Others)

Skills: L (Listening), S (Speaking), R (Reading), W (Writing), I (Integrated), G (Grammar), V (Vocabulary)Student Modality: I (Individual), P (Pair), C (Choral), G (Group)

Student modality: Individual (I), Pair (P), Group (G), Choral (C)

2014 September TRACE (PRE-TEST) Item Analysis

Total Possible Points: 65	Median Score: 53,4	Maximum Score: 49
Total Students: 147	Mean Score: 52,5	Minimum Score: 22
Standard Deviation:5,74	TRACE Reliability Coefficient (KR20): .60 TRACE Listening: .40 TRACE Reading: .54	Range of Scores: 29

No	Question	Correct Answer	Response Frequency (%)			Item Difficulty (%)			Item Discrimination		
			A	B	C	A	B	C	Lower 27%	Upper 27 %	Point Biserial
1	R1	3	3	23	73	3	23	73	27	31	0,02
2	R2	3	10	3	86	10	3	86	34	34	0,00
3	R3	2	12	62	26	12	62	26	19	32	0,09
4	R4	1	41	50	8	41	50	8	16	21	0,03
5	R5	3	23	29	47	23	29	47	17	22	0,03
6	R6	2	23	51	26	23	51	26	20	21	0,01
7	R7	1	54	15	31	54	15	31	17	24	0,05
8	R8	2	36	59	5	36	59	5	18	31	0,09
9	R9	3	27	45	28	27	45	28	14	19	0,03
10	R10	2	15	59	26	15	59	26	21	26	0,03
11	R11	3	69	21	10	69	21	10	5	6	0,01
12	R12	2	20	61	20	20	61	20	17	29	0,08
13	R13	1	91	7	2	91	7	2	34	39	0,03
14	R14	1	72	21	7	72	21	7	21	35	0,10
15	R15	3	4	50	46	4	50	46	14	23	0,06
16	R16	2	5	35	60	5	35	60	8	16	0,05
17	R17	2	5	89	5	5	89	5	30	39	0,06
18	R18	1	35	9	56	35	9	56	5	27	0,15
19	R19	1	59	8	32	59	8	32	22	24	0,01
20	R20	1	45	55		45	55		6	29	0,20
21	R21	1	73	27		73	27		23	37	0,10
22	R22	2	69	31		69	31		22	37	0,10
23	R23	2	80	20		80	20		22	37	0,10
24	R24	2	41	59		41	59		12	22	0,07
25	R25	2	43	57		43	57		12	24	0,08
26	R26	1	75	11	14	75	11	14	20	35	0,10
27	R27	3	1	22	77	1	22	77	29	33	0,03
28	R28	1	67	29	4	67	29	4	21	28	0,05

29	R29	2	1	88	11	1	88	11	30	37	0,05
30	R30	2	2	77	21	2	77	21	29	31	0,01
31	R31	3	18	6	76	18	6	76	28	35	0,05
32	R32	1	35	42	23	35	42	23	9	12	0,02
33	R33	1	70	9	21	70	9	21	25	30	0,03
34	R34	1	48	26	26	48	26	26	15	25	0,07
35	R35	2	16	61	23	16	61	23	18	28	0,07
36	R36	3	23	11	66	23	11	66	18	34	0,11
37	R37	2	8	55	37	8	55	37	10	29	0,13
38	R38	3	6	10	83	6	10	83	24	36	0,08
39	R39	2	12	86	3	12	86	3	29	36	0,05
40	R40	1	91	5	4	91	5	4	36	39	0,02
41	L1	2	39	41	20	39	41	20	10	17	0,05
42	L2	3	7	27	66	7	27	66	26	31	0,03
43	L3	2	38	24	38	38	24	38	4	15	0,08
44	L4	2	18	42	40	18	42	40	7	21	0,10
45	L5	3	58	23	19	58	23	19	4	7	0,02
46	L6	1	53	22	25	53	22	25	22	24	0,01
47	L7	2	26	20	54	26	20	54	4	13	0,06
48	L8	3	13	35	52	13	35	52	15	28	0,09
49	L9	1	81	16	3	81	16	3	27	37	0,07
50	L10	2	58	33	10	58	33	10	11	13	0,01
51	L11	1	86	12	2	86	12	2	29	38	0,03
52	L12	3	10	51	39	10	51	39	13	18	0,03
53	L13	3	35	10	54	35	10	54	17	23	0,04
54	L14	1	57	18	26	57	18	26	19	21	0,01
55	L15	2	3	84	13	3	84	13	32	34	0,01
56	L16	3	18	23	59	18	23	59	22	26	0,03
57	L17	3	25	37	37	25	37	37	11	16	0,03
58	L18	3	53	33	14	53	33	14	7	5	-0,01
59	L19	1	45	13	42	45	13	42	9	25	0,11
60	L20	2	6	65	29	6	65	29	13	33	0,14
61	L21	2	28	50	23	28	50	23	16	26	0,10
62	L22	3	24	23	53	24	23	23	14	24	0,10
63	L23	2	17	71	12	17	71	12	21	31	0,10
64	L24	1	43	26	31	43	26	31	12	22	0,10
65	L25	1	55	16	28	55	16	28	14	24	0,10

APPENDIX I

2015 January TRACE (POST-TEST) Item Analysis

Total Possible Points: 65	Median Score: 74,5	Maximum Score: 62
Total Students: 147	Mean Score: 73,98	Minimum Score: 39
Standard Deviation: 6.30	TRACE Reliability Coefficient (KR20): .62 TRACE Listening: .44 TRACE Reading: .50	Range of Scores: 40.6

No	Question	Correct Answer	Response Frequency (%)			Item Difficulty (%)			Item Discrimination		
			A	B	C	A	B	C	Lower 27%	Upper 27 %	Point Biserial
1	R1	3	33	28	39	33	28	39	38	50	0,03
2	R2	1	84	8	7	84	8	7	80	93	0,03
3	R3	1	90	5	5	90	5	5	80	95	0,04
4	R4	3	41	6	52	41	6	52	48	65	0,05
5	R5	3	18	10	71	18	10	71	45	88	0,12
6	R6	2	0	100	0	0	100	0	40	40	0,00
7	R7	3	2	20	78	2	20	78	28	34	0,04
8	R8	3	18	3	79	18	3	79	27	38	0,06
9	R9	3	16	25	60	16	25	60	23	26	0,02
10	R10	1	66	24	10	66	24	10	26	29	0,02
11	R11	1	88	12	0	88	12	0	33	39	0,04
12	R12	2	2	97	1	2	97	1	36	40	0,03
13	R13	3	3	10	86	3	10	86	33	35	0,01
14	R14	1	74	7	18	74	7	18	21	37	0,11
15	R15	3	0	1	99	0	1	99	38	40	0,01
16	R16	2	5	88	6	5	88	6	31	38	0,05
17	R17	3	1	2	97	1	2	97	37	40	0,02
18	R18	2	17	69	14	17	69	14	29	30	0,01
19	R19	2	25	71	4	25	71		22	32	0,07
20	R20	2	86	14		86	14		30	38	0,05
21	R21	2	82	18		82	18		27	36	0,06
22	R22	2	80	20		80	20		29	38	0,06
23	R23	2	49	51		49	51		16	26	0,07
24	R24	2	88	12		88	12		34	34	0,00
25	R25	2	80	20		80	20		30	34	0,03
26	R26	2	8	75	16	8	75	16	25	36	0,08
27	R27	3	6	32	62	6	32	62	15	36	0,10
28	R28	3	20	3	77	20	3	77	22	37	0,10
29	R29	3	2	18	80	2	18	80	29	39	0,07

240

30	R30	3	15	60	25	15	60	25	8	14	0,04
31	R31	2	3	50	47	3	50	47	10	27	0,12
32	R32	1	93	0	7	93	0	7	35	39	0,03
33	R33	1	87	5	7	87	5	7	30	37	0,05
34	R34	3	7	13	80	7	13	80	23	37	0,10
35	R35	1	82	16	2	82	16	2	26	37	0,80
36	R36	3	0	5	95	0	5	95	35	40	0,03
37	R37	1	41	36	23	41	36	23	12	21	0,06
38	R38	2	22	67	11	22	67	11	18	34	0,11
39	R39	2	8	86	5	8	86	5	29	40	0,08
40	R40	2	0	96	4	0	96	4	36	40	0,03
41	L1	3	5	16	78	5	16	78	27	36	0,06
42	L2	1	78	16	6	78	16	6	23	34	0,08
43	L3	2	25	69	6	25	69	6	21	31	0,07
44	L4	1	69	31	1	69	31	1	22	32	0,07
45	L5	2	1	86	13	1	86	13	29	37	0,05
46	L6	3	1	2	97	1	2	97	38	39	0,01
47	L7	2	7	86	6	7	86	6	28	38	0,07
48	L8	2	2	93	5	2	93	5	35	39	0,03
49	L9	2	19	50	31	19	50	31	13	27	0,10
50	L10	1	54	20	25	54	20	25	14	30	0,11
51	L11	1	63	29	8	63	29	8	24	26	0,01
52	L12	2	4	95	1	4	95	1	36	40	0,03
53	L13	3	10	9	82	10	9	82	27	37	0,07
54	L14	3	5	1	95	5	1	95	36	39	0,02
55	L15	3	4	1	95	4	1	95	36	39	0,02
56	L16	1	83	16	1	83	16	1	29	36	0,05
57	L17	2	5	93	2	5	93	2	37	38	0,01
58	L18	3	2	7	91	2	7	91	34	39	0,03
59	L19	1	83	4	13	83	4	13	30	38	0,05
60	L20	2	12	85	3	12	85	3	31	38	0,05
61	L21	2	1	95	4	1	95	5	37	39	0,01
62	L22	2	4	2	94	4	2	94	37	38	0,01
63	L23	3	3	3	94	3	3	94	36	39	0,02
64	L24	1	94	6	0	94	6	0	36	39	0,02
65	L25	3	3	2	95	3	2	95	36	39	0,02