Advances in Language and Literary Studies
ISSN: 2203-4714
www.alls.aiac.org.au

Can Exams Change How and What Learners Learn? Investigating the Washback Effect of a University English Language Proficiency Test in the Turkish Context

Asli Lidice Gokturk Saglam[1]*, Hossein Farhady[2]

[1]*Özyeğin University, School of Languages Çekmeköy Campus Nişantepe District, Orman Street, 34794 Çekmeköy, İstanbul, Turkey*
[2]*English Language Teaching Program, Yeditepe University, İnönü Mah, Kayışdağı Cd. 326A, 34755 Maltepe, İstanbul, Turkey*

Corresponding Author: Asli Lidice Gokturk Saglam, E-mail: asli.saglam@ozyegin.edu.tr

ARTICLE INFO

Article history
Received: October 10, 2018
Accepted: January 25, 2019
Published: February 28, 2019
Volume: 10 Issue: 1
Advance access: January 2019

Conflicts of interest: None
Funding: None

Key words:
Washback Effect,
Theme-Based Language Proficiency Test,
Integrated Language Proficiency Test,
Washback on Learning,
English for Academic Purposes

ABSTRACT

This article reports on a mixed-method study that examined the washback of a local integrated theme-based high-stakes English language proficiency test that is used in a university English for Academic Purposes (EAP) program in Turkey. The assumption behind employing an integrated theme-based test, which resembles authentic language use, was that it would bring about a positive washback on learning (Leki, Cumming & Silva, 2008: Leki & Carson 1997). The data were collected from both focus-group interviews after the instruction and pre- and post- proficiency test scores of 147 EFL students in the Preparatory English Language Program (PEP). Test of Readiness for Academic English (TRACE) was administered at the beginning and at the end of a 4-month English language instruction period. Repeated measure ANOVA and inductive analysis of the transcribed interview data were used for analyzing quantitative and qualitative data respectively. The findings indicated that the test had both positive and negative washback on the learning. Most students considered that using source-based information and their notes taken during the listening task into their writing raised their awareness in terms of generating, organizing and linking ideas as well as modelling vocabulary and sentence structures. However, the test also exerted negative washback upon learning since students were inclined to prioritize test-oriented practice. The implications of the study suggest that a theme-based integrated proficiency exam may elicit positive washback on learning that could be used for validity evidence in EAP contexts and lead to more appropriate language assessment. The procedures are detailed, the findings are presented and discussed, the applications and implications for teachers and test designers are explained, and some suggestions are made for further research.

INTRODUCTION

The influence of tests on teaching and learning has been conceptualized as "washback" (Alderson & Wall, 1993), that can be both positive and negative. Washback models (Alderson & Wall, 1993; Bailey, 1996; Hughes,1994) offer guidelines for measuring its magnitude and pursuing positive test influence. In test-oriented contexts such as Turkey, high-stakes tests have been reported to exert strong negative washback on teaching and learning (Akpinar & Cakıldere, 2013; Karabulut, 2007; Ozmen, 2011; Sevimli, 2007). Within such test-dominant contexts, utilizing tests as a lever to engineer positive washback in education gains importance.

In Turkish context, the official framework for assessing a second language ability often requires measuring all four skills (speaking, listening, reading, and writing) as independent constructs. However, such a framework may fall short in measuring language proficiency in authentic communicative contexts since skills are used in an integrated way in the real world (Hillocks, 2002; Sawaki, Quinlan & Lee, 2013). The issue becomes more significant within the domain of English for Academic Purposes (EAP), since academic studies rely on integrating two or more of language skills. Thus, utilizing a theme-based integrated skills proficiency test is grounded in complying with authenticity arguments (Gebril, 2009; Plakans, 2009; Weigle, 2004). Similarly, for Cumming (2013) the construct of integrated assessment mirrors academic literacy activities since in many academic contexts, writing requires the integration of reading and listening. Additionally, reading-to-writing and listening-to-writing tasks "provide content, and, thus increase equity by minimizing the impact on writing performances of background knowledge, creativity, and life experience" (Plakans & Gebril, 2017, p. 98).

Given the desirable role of integrated skills assessment on student performance, researchers call for more detailed examination of its washback effect in different instructional contexts (Delaney, 2008; Plakans, 2009; Plakans & Gebril, 2013; Weigle, Yang, & Montee, 2013: Yu, 2013). It is often accentuated that when there is a curricular alignment in a language program between what is taught and what is tested, washback is apt to be strong (Madaus, 1988; Smith, 1991) since "what is assessed becomes what is valued, which

Published by Australian International Academic Centre PTY.LTD.

http://dx.doi.org/10.7575/aiac.alls.v.10n.1p.177

becomes what is taught" (McEwen, 1995 in Cheng & Curtis, 2004, p. 3). It is also argued that this alignment between what is taught and what is tested on the one hand, and what the real life demands of academia are, on the other, foster student motivation and facilitate transfer of language skills to academic courses (Leki & Carson,1994, 1997 in Plakans & Gebril, 2013). Thus, due to authenticity arguments and its facilitating effect on learning, integrated assessment may be efficient in engineering positive washback.

The research context of this study is a proficiency testing situation at the tertiary level in a Turkish university. The Test of Readiness of Academic English (TRACE) aims to assess whether language learners have the sufficient ability to use English for academic purposes in university classrooms. In an effort to replicate tasks that students encounter in their classes, TRACE entails four readings and a lecture listening which acts as source materials for the test takers in a source-based writing task, in the form of an extended essay. TRACE functions in a local context in which such alignment is pursued through teaching activities and assessment tasks that replicate target language use (TLU) domain. An exam which assesses language proficiency via authentic theme-based integrated assessment tasks may be more valid since it mirrors the constructs of real-life demands of academic life. The aim of this research study is to examine the validity of the assumption that employing an integrated theme-based test of English language proficiency that is similar to authentic language use in the tertiary education context in Turkey would bring about positive test influence upon learning. Therefore, investigating its washback is of vital importance for engineering positive washback.

The Concept of Washback

Washback is conceptualized as the influence of tests on teaching and learning. According to Bailey (1996), washback is a multifaceted phenomenon that involves many factors including participants (students, teachers, administrators, materials writers and publishers), processes (materials development, syllabus design, modifications in instruction and methodology, use of learning and/or test taking strategies), and products (learners' intake, skills and quality of learning). Furthermore, it is believed that some effects of a test may be beneficial for the development of the learners' progress and achievement, whereas others may be detrimental (Alderson and Wall 1993, Brown and Hudson 2002, Hughes, 2003). Consequently, some scholars have suggested using strategies that would lead to positive washback to engineer curricular and instructional changes as well as increasing learner innovations. Similarly, it's often asserted that the quickest and most effective way to change student learning is to change the methods of assessment (Brown,1997; Elton & Laurillard,1979). Thus, the use of assessment as a lever for a curricular change is widely accepted not only in education (Chapman & Synder, 2000; James, 2000) but also in language education (Cheng, 1997, 1998; Pearson, 1988; Swain, 1985; Wall and Alderson, 1993).

One of the influential factors in almost all models of washback is reported to be learners' use of strategies. However, research into washback has led to some contradictory findings in the use of (Watanabe, 1992) or the influence of learning strategies (Gosa, 2012; Pan & Newfields, 2012; Zhan and Andrews, 2013) in and out of class contexts. Pan and Newfields (2012) explored the effect of mandated EFL proficiency tests on learners in tertiary level institutions with and without English language proficiency requirement. They concluded 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. They also stated that test requirements did not lead to "studying for the test," that is often reported in examination-oriented educational settings (Chern, 2002; Lai, 2003; Tsai & Tsou, 2009 in Pan & Newfields, 2012 p. 119). Interestingly, their findings contradict the conclusions of some other washback researchers (Green 2007, 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 and non-communicative approaches, a change in students' learning activities was not observed.

These washback studies are informative with respect to effects of examinations on learning. However, research into the washback of tests on learners remains limited. Zhan and Andrews (2013) concluded that students were more likely to change what they learned rather than how they learned. Findings of their study resonate with the conclusions of previous washback studies (Bailey, 1996; Ferman, 2004; Green 2007; Shih, 2007) in that students attach importance to skills that are tested. Therefore, they often study for the test without adopting changes in their learning strategies. 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. The emphasis on the student perception of washback in the field (Zhan & Andrews, 2013) has prompted the present study to consider individual learners and their understanding toward how the test affects their learning.

Although one of the early washback studies was conducted in Turkey by Hughes (1988), there are few studies on the effects of nationwide tests on student learning in the Turkish educational context. Contrary to Hughes' findings, these studies reported that there was negative washback on learning. Ozmen (2011) examined the washback effect of Intercollegiate Foreign Language Examination (ÜDS) on candidates for academic positions in Turkey. Their findings revealed that participants needed to develop more than what the test assessed and therefore, the test was considered as an obstacle for their learning. Akpinar and Cakıldere (2013) indicated that tests exerted positive impact on the tested skill of reading, but negative impact on other skills of writing, listening and speaking since they are not tested. It was concluded that participants were highly interested in improving their reading skills to get higher scores at the expense of ignoring other skills.

Most of the above-mentioned studies addressed the washback of normal traditional tests. However, to the best

of our knowledge, no empirical study has been reported thus far that explored test effects of a theme-based integrated language proficiency test in an EAP setting. Thus, this study aims to fulfil this research gap and contribute to the development of washback related studies by using an integrated assessment tasks. It also attempts to explore another dimension of research on washback by using an integrated skills tasks since no clear evidence is available on its effect on learning. Therefore, the present study delves into student perceptions regarding test effects and gains in scores by formulating the following questions:

1. Is there a potential washback effect of the TRACE on learning?
2. Does the language instruction program, based on EAP skills, lead to gains in scores on the writing, listening, and reading parts of the TRACE?

METHOD

The Research Context

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 language ability and academic skills to meet the language requirements of their major fields of study. The levels of PEP are aligned with those offered by the Common European Language Framework, namely elementary (A1), pre-intermediate (A2), intermediate (B1), upper-intermediate (B2) and advanced level (B2+). Based on the scores of the placement test, incoming students at A1, A2, and B1 levels in English are directed to intensive general English courses in PEP programs. Others are required to take the TRACE which is an institutional proficiency test. The main purpose of TRACE is to determine whether the test-taker's skills and language level are sufficient for academic coursework. Thus, 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. Those who score between 50-65 are placed in advanced level and those scoring lower than 50 are directed to upper-intermediate level in PEP. In other words, TRACE functions as both the proficiency and placement test.

Design of the Study

This study adopted a mixed method design to ensure internal validity. As Creswell and Plano Clark (2006) indicate, mixed methods are used to enrich the findings of a single approach. Fielding and Fielding (2008) describe this purpose of mixed-method research designs as "complementary" since it provides triangulation with the aim of convergence, corroboration and correspondence of results from different methods (p. 558). Similarly, washback researchers advocate employing a mixed-method approach and using multiple sources of data. 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. Therefore, this study integrated both qualitative and quantitative data obtained by different instruments including classroom observations, questionnaires, interviews and pre- and post- proficiency scores to ensure a more valid interpretation of the findings.

Participants

Participants were 147 incoming students to the PEP who were placed at upper-intermediate (n=44) and advanced level (n=103) at the onset of 2014-2015 academic year based on their TRACE scores. The TRACE was administered as pre-test at the beginning of a 4-month English language instruction period and as a post-test at the end of instruction. Participants with similar proficiency levels were selected based on their pre-test scores. Upper and advanced level students received 4 hours of instruction per day totaling 320 hours for the whole 16-weeks semester. The focus of the instruction was to improve students' language skills directed towards EAP. Participants were exposed to these language skills in an integrated way through in-house supplementary materials which were aligned with the proficiency exam. The students, aged between 18 and 23, came from different cities of Turkey and they had diverse educational background. However, since they were required to take multiple-choice tests for admission to high schools and then to university before they come to PEP, it was assumed that they were familiar with gatekeeper high-stakes exams, exam preparation, and multiple-choice exam format.

Data collection and analysis

Data were collected through a variety of instruments using both qualitative and quantitative procedures. The data included information from focus-group student interviews as well as pre- and post- language proficiency test scores of 147 EFL students who were enrolled in the PEP.

TRACE

The main instrument used in this study was the Test of Readiness for Academic English (TRACE) which is an institutional English language proficiency test. It has adopted a theme-based and integrated skills approach in an effort to reflect the actual language use in academic domains as closely as possible. Although English language proficiency assessment has long depended on discrete testing, more recently theme-based exams, which utilize reading-to writing and listening to writing on a given theme, have been adopted since they are perceived to provide an authentic representation of language use in an academic context. Examples are Canadian Academic English Language Assessment (CAEL); the English proficiency exam of the Universidad Veracruzana (EXAVER) in Mexico; and English Proficiency Exam (EPE) of Middle East Technical University in Turkey.

There are four sections in the TRACE; introduction, reading, listening, and writing. All sections are on a single theme. They are usually selected from the field of psychology, sociology, environment or business that test takers

would be familiar with. In the introduction, test-takers are exposed to visuals and required to brainstorm about the topic and take notes on a note-taking sheet. Then, they read multiple texts about the same topic and respond to multiple choice comprehension questions. Reading comprehension items also require cross-textual reference for which test-takers need to consider all the readings in the test. The third section includes listening to a lecture and taking notes. The final section entails an essay writing task using a variety of sources (ideas from readings, lecture notes from the listening and notes from introduction section). Integration of skills encourages authenticity since academic writing tasks in EAP courses and university content courses commonly resort to use of external sources (Leki & Carson, 1994, 1997). Adhering to authenticity argument, test developers integrate three language skills (reading, listening and writing) in TRACE to reflect the target language use in academic settings.

TRACE is used in a local context where curricular alignment is attempted through teaching activities and assessment tasks that resemble real life academic activities. A major assumption underlying this study was the assumption that employing an integrated theme-based test of English language proficiency that approximates authentic language use would bring about positive washback on teaching and learning.

Interviews

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. Some of the interviews were done in English whereas some of them were conducted in Turkish based on the preference of the students. Focus-group interviews took between 20-35 minutes. Interviews were performed with three purposes in mind. First, they were expected to encourage active group interaction (Barbour, 2007, p. 2) which could provide insights to students' perceptions of teaching materials, classroom activities and the extent of their learning. Second, interview questions were designed to elicit information on students' self-evaluation of reading, listening and writing competency. Third, interviews aimed to provide information on students' attitudes towards teaching materials, tasks and correspondence between teaching-learning activities that would lead to success on TRACE.

The semi-structured interviews were transcribed and analyzed using Bogdan and Biklen's (1998) framework. Through careful examination of the transcripts, conceptual themes were identified adhering to recurring words and ideas. The emerging conceptual categories which led to major themes, were classified and used to support research findings. Additionally, the results were quantified where possible to get a preliminary overview of data. During analysis of transcripts of interview data, two academicians who held PhD degrees in ELT were consulted to benchmark the meaning coding, condensing the meaning, and interpreting the outcome. Coding schemes were compared to identify similarities and differences. To determine the interrater reliability, number of agreements was divided by total number of agreements and disagreements. The disagreements were resolved in further meetings. Interview questions were piloted with teachers and students. Also, following Qi (2004), tape recordings, field notes, codes, and analysis sheets were kept as audit trail.

FINDINGS

Analysis of Qualitative Data

In the first part of the focus-group interviews students were asked about their progress in the advanced and upper-intermediate PEP courses. In the second part, they were asked questions to elicit their ideas regarding the washback of TRACE on teaching materials and the methodology employed by their teachers. The majority of 47 respondents claimed that they had improved their language ability. Table 1 summarizes their responses.

The findings support the data obtained from different administrations of TRACE. However, the progress is more observable in writing and listening skills. Some pointed out that small progress in reading was due to the exclusion of explicit skills training. In addition, improvement in speaking skill received the lowest rating from the students. Students attributed it to limited time allocation to speaking activities at the cost of narrowing curriculum and class instruction towards tested skills. It was also claimed that some students prioritized focus on tested skills rather than critical reading or meaningful learning.

Further analysis of interview transcripts revealed three categories of potentially influential factors on the TRACE washback. These categories included exam-orientedness factors, materials induced factors, and teacher induced factors. Table 2 presents the frequency of these categories.

Table 1. Student perceptions on improving their language ability

	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
Speaking skills	20	15	12	43	32	26
Grammar	47	0	0	100	0	0
Vocabulary	47	0	0	100	0	0

Table 2. Frequency of emerging themes and sub-themes of student interviews

	f	%
Exam orientedness	31	74
Materials induced factors		
Synthesis of information from sources and integration of skills	19	40
Teacher Induced factors		
Effect of teacher's practice- Variation between teachers	22	47
Coaching	16	55

Exam orientedness

The findings indicated that the TRACE had a negative washback on student learning since students were inclined to favour activities intended for test orientation and coaching. Due to their test-oriented background, many respondents concurred that they got used to multiple choice high stakes testing culture which requires significantly different abilities in comparison to academic demands of the PEP and the university. However, as a proficiency test, TRACE was similar to the tests they had experienced before. Consequently, students claimed that they consider the requirements of the exam to be at the center of their learning process and they often focus on test taking strategies. One of the students claimed; *"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".* It can be inferred that although previous educational background reinforced rote learning and had not fostered communicative and creative language learning through authentic materials, this theme-based integrated approach to testing brought about new skills such as 'critical thinking' to student learning processes.

In addition, students were asked to self-evaluate themselves and share their perceptions regarding their improvement in the courses. There were comments focusing on strategies to get better grades on the test rather than critical self-evaluation of their learning. Their responses indicated that most were highly exam-oriented and lacked awareness on how they progressed in the PEP. Instead of approaching efficient learning, as seen in the example quotation below, some respondents claimed not to pay attention to learning a foreign language but learning certain skills, excluding speaking which is not tested on the exam. *"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 the majority of the students tended to focus on test taking skills and strategies for increasing their scores rather than a deliberate focus on learning.

Another set of responses referred to the significance of memorizing complex sentences to use in essays with the intention of getting higher scores. For example, one participant said, "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 important to note 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 improving language ability. The formula of getting high scores in exams was prescribed as memorizing newly learned vocabulary items as well as grammar structures and using them in the writing part of the exam. These findings support the significant differences between mean scores of different administrations of TRACE (pre- and post-scores) which is interpreted as an indicator of washback on their learning.

In a similar line of thought, when they were 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 indicate a negative washback effect on student learning since students were inclined to prioritize test-oriented practice. They claimed, as previous findings reported (Gosa, 2004; Shih, 2007, Tsagari, 2009), that test oriented activities and test-specific coaching as 'the most beneficial' to prepare for the high-stakes proficiency exam.

Materials related factors

Majority of the students (72%) showed negative perceptions towards the course books 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. These mismatches included 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 supplementary multiple-choice test tasks.

The findings resonate with the conclusions of previous washback studies (Bailey, 1996; Ferman, 2004; Shih, 2007;

Green 2007) in that students attach importance to skills that are similar to those in the test (Zhan & Andrews, 2013). 40% of the respondents indicated that TRACE impacted materials, especially those in-house prepared supplementary materials. The reason was that task types, question types, genres and length of the texts resembled those in the exam.

One respondent commented*: "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"*.

Students' responses also indicate that synthesis of source-based information had a facilitative effect on their learning when they had good language competency as stated in the following comment: *"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"*.

It was also mentioned that prior classroom practice on synthesizing information from a variety of sources 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 test because it raised students' awareness to generating, organizing and linking ideas as well as modelling vocabulary and sentence structures. It was frequently mentioned that supplementary materials, which were aligned with the test, encouraged students to synthesize ideas across a variety of sources and had a positive effect on students' learning.

Teacher induced factors

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

Effect of teacher's practice- Variation between teachers

47% of the student response indicated that teachers' classroom activities involved variety regarding the use of different media and modality. A participant stated: "*We don't only do listening exercises. We also watch videos to understand the topic better. This also helps our organization in writing because we get different ideas. Our teacher gives us links to websites for practicing 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"*. 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, whereas in some others, there was no sign of an integrated approach or utilization of different types of materials. These findings implied that teachers' knowledge of the nature of the exam, influenced the variety of instructional strategies that would help students in the test.

Coaching

For more than half of the students (55%), especially for those at the advanced level, the focus of the instruction in the classroom was in line with exam practice. Some students even distinguished between advanced and upper-intermediate level courses. They claimed that the former was more of a course in the direction of preparing them for the TRACE 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"*.

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 the content and language from different classroom sources. 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.

Analysis of Quantitative Data

The quantitative data was collected from 147 participants selected form over 800 incoming students. 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. The low reliability might be attributed to the shrinkage of variance in the scores of the participants due to their close levels of language ability. To gain insight into possible relationships between the overall scores and those of other sections, correlation analysis was carried out and presented in Table 3.

These correlation coefficients do show a good go togetherness of students' performance on different components of the test. More importantly, there are low intra-correlations among the scores of different sections of the test, as well as inter-correlations among the different administrations of the tests. Other than the variance shrinkage due to truncated data, most of the correlations, that are well below the normally expected values, may lead to the conclusion that the test does not enjoy high reliability and validity in the PEP.

To investigate the differences among the scores of students on different administrations of the TRACE, a repeated measure ANOVA was used to compare the mean scores

Table 3. Correlation analysis of scores

	Reading September	Listening September	Writing September	Overall grade September
Reading September	1	0.25**	–0.22**	0.55**
Listening September	0.25**	1	–0.07	0.69**
	0.00		0.38	0.00
Writing September	–0.22**	–0.07	1	0.46**
	0.01	0.38		0.00
Reading January	0.25**	0.14	–0.20*	0.09
	0.00	0.09	0.02	0.27
Listening January	0.12	0.13	–0.13	0.06
	0.14	0.13	0.12	0.46
Writing January	–0.15	–0.14	0.19*	–0.05
	0.08	0.08	0.02	0.57
Overall grade Jan	0.06	0.02	–0.02	0.04
	0.46	0.80	0.83	0.68

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

Table 4. 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

of the students in reading, listening and writing as well as overall scores on TRACE that was administered in September 2014 and in January 2015. The scores belonged to the same students at two ability levels before and after the 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, p=.00 and for January test scores S-W=.97, df=147, p=.00). ANOVA results showed significant differences between pre and post test scores as an indication of washback of TRACE on learning (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 also demonstrated that there was a significant effect of instruction on overall gain scores, F(1. 146) = 969,45, p=.000. This indicates 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 4 below.

DISCUSSION AND CONCLUSION

In relation to the first research question which attempted at exploring potential washback of the TRACE, student interviews and pre- and post-test scores indicated that there could be both positive and negative test effects exerted on the choice of materials, classroom activities, and learning outcomes. This study found that TRACE could lead to negative washback in the form of learning strategies geared towards being successful on the test and narrowing of learning towards tested skills. Students indicated that they often adopted a narrow scope of learning by overemphasizing their preference of test taking strategies as a learning strategy in order to boost their test scores. 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 (Shih, 2007, Tsagari, 2009; Gosa, 2004).

However, there was signs of positive washback on choice of materials intended to boost the scores through using integrated skills tasks. Probably more important, as a response to the second research question the proficiency scores obtained before and after the instruction of PEP revealed that TRACE was sensitive to instruction. The findings show that the language ability of the students show significant improvement on the three language skills. This finding was also supported by the data obtained through the student interviews.

Findings of the qualitative data also supported both negative and positive washback of TRACE. Students were inclined to be test-oriented because they claimed to value activities that were oriented towards the test or test-specific coaching. The implications of the study suggest that a theme-based integrated proficiency exam may elicit positive washback on learning that could be used for validity evidence in EAP contexts and lead to more appropriate language assessment.

Like many studies, this study is not without limitations either. 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 the generalizability of the

findings to the broader English language teaching and testing populations in other contexts could not be appropriate. Some researchers (e.g. Perrin, 2000; Tsagari, 2006) argued that any washback research is innately context-based. However, investigating those variables in a specific educational context may hopefully shed some light on similar variables in similar contexts. Findings may also have implications for EFL students, 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 i.e. TRACE exam scores. Since the reliability indexes for both pre- and post-test scores as well as the correlations between sections of the exam were, the test may require some modifications.

Further research may address the limitation of this study and open a new line of research by relating learners' test scores to their perceptions and their real performance. Finally, Exploring the link between individual learner's test scores and their perceptions may bring about insights in to washback research.

REFERENCES

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.

Amrein, A.L. & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18), 1-74.

Bailey, K.M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(3), 257-279. https://doi.org/10.1177/026553229601300303

Becker, B. (1990). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research. 60*(3), 373-417. https://doi.org/10.3102/00346543060003373

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly, 10*(1), 1–8. https://doi.org/10.1080/15434303.2011.622016

Green, A. (2007). Washback to learning outcomes: a comparative study of IELTS preparation and university professional language courses. *Assessment in Education, 14*(1), 75-97. https://doi.org/10.1080/09695940701272880

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.

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.

Hillocks, G. (2002). *The testing trap: How state assessments of writing control learning,* New York, NY: Teachers College Press.

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.

Klein, S.P., Hamiltion, L.S., Mc Caffrey, D. F. & Stecher, B. M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives,* 8, 49. http://epaa.asu.edu/epaa/v8n49

Leki, I., Carson, J. (1997). Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly, 31(1),* 36–69. https://doi.org/10.2307/3587974

Leki, I., Carson, J., &Silva, T. (2008*). A synthesis of research on second language writing*. London, UK: Routledge.

Özmen, K. (2011). Washback effects of the inter-university foreign language examination on foreign language competences of candidate academics. *Novitas-ROYAL Research on Youth and Language, 5 (*2), p. 215-228.

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.

Plakans, L. (2009a). Discourse synthesis in integrated second language writing assessment. *Language Testing, 26*(4), 561–587. https://doi.org/10.1177/0265532209340192

Plakans, L. (2009b). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes, 8*(4), 252–266. https://doi.org/10.1016/j.jeap.2009.05.001

Plakans, L., Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing, 17*(1), 18–34. https://doi.org/10.1016/j.asw.2011.09.002

Qi, L. (2004). Has a high-stakes test produced the intended changes. In L. Cheng, Y. Watanabe, & A. Curtis (Ed.), *Washback in language testing: Research contexts and methods* (pp. 171-191). Manwah, New Jersey: Lawrence Erlbaum Associates.

Robb, T. N., & Ercanbrack, J. (1999). A study of the effect of direct test preparation on the TOEIC scores of Japanese university students. *Teaching English as a Second or Foreign Language, 3*(4), 2-16

Read, J. &Hayes, B. (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.

Shih, C. (2007). A new washback model of students' learning. *The Canadian Modern Language Review, 64*(1)*,* 135-162. http://dx.doi.org/10.3138/cmlr.64.1.135

Shohamy, E., Donitsa-Schmidt, S. & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13(3)*, 298-317. https://doi.org/10.1177/026553229601300305

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. http://dx.doi.org/10.1080/15434303.2011.633305

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). https://eric.ed.gov/?id=ED497709 http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED497709&ERICExtSearch_SearchType_0=eric_accno&accno=ED497709

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. https://doi.org/10.1080/15434303.2012.750660

Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly, 10*(2), 196-218. https://doi.org/10.1080/15434303.2012.721423

Xie, Q. & Andrews, S (2012). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modelling. *Language Testing*,*30*(1), 1-22. https://doi.org/10.1177/0265532212442634

Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly, 10*(1), 110–114. https://doi.org/10.1080/15434303.2013.766744

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. https://doi.org/10.1080/0969594X.2012.757546

APPENDIX A

Student Focus-Group Interview Topics and Questions

0. Opening & Introduction (Key points of the study, purpose, confidentiality, media and timing)
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?
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?
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?
7. Do you think that you improved yourself in speaking?
8. What do you think you have learned in terms of speaking skills? Can you give some examples?
2. Attitudes to Materials & Tasks
1. 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?
2. What kind(s) of reading, listening & writing activities and tasks have you done in the class?
3. Do you remember any task that was directly related to the test and it may help you improve your scores?
4. Do you think that content of the course (what you learned in class) and TRACE are similar? How?
5. How do you think what you learned in the course may help you in TRACE?
6. 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