

Learner Washback Variability in Standardized Exit Tests

August 2014 – Volume 18, Number 2

Yi-Ching Pan

National Pingtung University, Taiwan

<yichingpan@yahoo.com.tw>

Abstract

In much of the world, the issue of accountability and measurement of educational outcomes is highly controversial. Exit testing is part of the movement to ascertain what students have learned and hold institutions and teachers to account. However, compared to the large number of teacher washback studies, learner washback research is lacking (Cheng, 2008). By exploring standardized exit test washback in Taiwan, this study aims to test Alderson and Wall's hypotheses (1993) regarding learner washback variability, namely that a test will influence 1) degree/depth of learning, 2) attitudes toward methods of learning, and 3) some learners but not others. This study is one of very few washback studies that compare learner behaviors before and after the introduction of exit tests, thereby obtaining a clearer picture of washback effects on learners. Questionnaires and scores on two exit tests (GEPT and TOEIC) were collected from 589 Taiwanese university students before and after the introduction of English exit tests in Taiwan. This study highlights the differential effects of exit requirements for different groups of learners depending on their years of study, proficiency levels, and viewpoints of tests.

Keywords: washback effects, variability, English certification exit requirements

Introduction

In much of the world, the issue of accountability and measurement of educational outcomes is highly controversial. Exit testing is one part of the movement to ascertain what students have learned and to hold institutions and teachers to account (Jacob, 2001; O'Loughlin & Arkoudis, 2009; Humphreys and Mousavi, 2010). However, compared to the large number of teacher washback studies, learner washback research is lacking (Cheng 2008; Spratt, 2005; Wall, 2000; Watanabe, 2004). By exploring standardized exit test washback, this study tested Alderson and Wall's hypotheses (1993) regarding learner washback variability. In particular, this study aimed to investigate how exit tests influence learners' motivation, learning strategies, and test performance.

Theoretical Washback Framework

“Washback” (Alderson & Wall, 1993; Buck, 1988) or “backwash” (Biggs, 1995, 1996; Hughes, 2003) is often defined as the influence that tests exert on teaching and learning. To better understand both the scope and mechanism of washback, researchers have proposed various models, hypotheses, and concepts.

Hughes’s (1993) trichotomy of backwash model describes test effects in terms of “participants,” “process (es),” and “product(s).” A test could affect “participants” such as teachers, students, administrators, materials writers, and publishers, “all of whose perceptions and attitudes toward their work may be affected” (p. 2). “Process” refers to any actions undertaken by participants to complete teaching and learning tasks, such as materials development, syllabus design, changes in teaching methods or content, and learning and/or test-taking strategies (p. 2). “Product” refers to “learning outcomes (e.g., skills) and the quality of learning (fluency)” (p. 2). Hughes’s model indicates that how participants interpret and react toward a test affects how and what their response to it will be, implying that the quality of a test plays an important role in determining washback test effects. However, this model does not clearly delineate washback effects on various aspects relating to the teachers/learners and their teaching/learning processes.

Bailey’s (1996) model of washback, drawn from Hughes’s framework, specifically identifies Hughes’s participants and products. The participants include students, teachers, materials writers, and curriculum designers and researchers, whereas products refer to the test effects those participants will produce. This model emphasizes the possible interactions of these participants; however, it does not specify the possible differences in the impact of tests on individuals.

Alderson and Wall’s (1993, pp. 120-121) fifteen washback hypotheses address, in greater detail than Hughes’s and Bailey’s models, what washback effects might look like on teachers and learners, as stated below:

- A test will influence the rate and sequence of teaching/learning.
- A test will influence the degree and depth of teaching/learning.
- A test will influence attitudes about the content, method, etc., of teaching/learning.
- Tests that have important consequences will have washback.
- Tests that do not have important consequences will have no washback.
- Tests will have washback effects for some learners and some teachers but not for others.

Furthermore, these hypotheses imply that in addition to a test, there might be other factors that elicit more effects from some learners and teachers than they did for others. In addition, not only does the quality of a test affect teaching and learning, but also how it is used in different situations affects the strength of washback. However, no explanations have been made for several of their hypotheses regarding washback variability.

Cheng (1997, 2005), Watanabe (2004), and Green (2007) developed the concept of “washback intensity” to delineate possible factors that contribute to different forms of “washback variability”. In their viewpoints, participants’ perceptions of test stakes, test quality, and test difficulty tend to vary from person to person and therefore lead to differences in the washback effects experienced by individuals.

Since the theoretical studies discussed above do not involve evidence-based research, researchers such as Alderson & Hamp-Lyons (1996), Bailey (1999), Cheng (1997; 2010), and Wall (2005) cite the need for empirical studies to determine the relationship between how participants, processes, and products are affected by tests. In other words, washback variability and washback intensity must be empirically tested.

Empirical Washback Studies

So far the vast majority of foreign language washback studies—such as those by Alderson and Hamp-Lyons (1996), Burrows (2001, 2004), Chen (2002), Cheng (2004, 2005), Ferman (2004), Hawkey (2006), Lam (1994), Lewkowicz and Zawadowska-Kittel (2008), Qi (2005), Saif (1999, 2006), Valazza, (2008), Watanabe (1996a, 1996b, 2004), Wall and Horak (2006, 2008), and Wu (2008)—have focused on *teacher* perspectives concerning tests. Three types of claims have been made: (1) tests influence both *what* and *how* instruction occurs, (2) tests impact *what* is taught but not *how*, and (3) teacher response to test preparation and changes in tests varies from person to person based on both testing (e.g., viewpoint on the quality and content of the test) and mediating factors (e.g., teachers’ educational backgrounds, beliefs, concerns for students’ proficiency levels, and psychological factors).

In contrast to the wealth of teacher washback studies, learner washback research is severely lacking. Wall (2000, p. 502) laments the paucity of research and Cheng (2008), Spratt, (2005), and Watanabe (2004) emphasize the need to explore the impact of tests on learners, since they are directly affected by these tests. Studies that have focused on learning washback reveal varied and sometimes contradictory findings. Three types of student washback claims are documented.

First, tests influence *what* is learned, but not *how*. Cheng’s (1998, 2005), Qi’s (2004, 2005), and Stoneman’s (2006) studies revealed that the methods students used to prepare for different exams were similar. Traditional learning strategies and exam preparation methods such as reading textbooks, memorizing vocabulary and idioms, going through previous exams, or relying on test prep books were prevalent among their research subjects.

Second, tests have a variety of effects on students, varying according to their different viewpoints of the tests or their different levels of language proficiency. Stoneman (2006), Tsagari (2006), and Shohamy et al. (1996) discovered that learners’ perceptions of the stakes and status of the test influenced the strength of the test effects: a high-stakes, high-status test promoted learning; students spent more time engaging in learning language skills that were covered on the test than they did on lower-stakes or lower-status tests. Similarly, Tsai and Tsou (2009) found

that negative student opinions on the adoption of standardized exit tests led to a decrease in motivation to learn English because their classes were test-oriented, only enhancing their test-taking skills instead of their communicative competence.

In regard to learner washback from language proficiency, the findings were mixed but appear to contend that students' levels of English proficiency carried some weight in determining the extent of the effort they were likely to make toward a test. According to Stoneman (2006), Watanabe (2001), and Chu (2009), low-achieving students tended to be more worried about the test or test requirement than high-achieving students, and these low-achieving students did not prepare for the test until the last minute or did not prepare at all. As contended by Watanabe, a test of appropriate difficulty for the learner can positively affect his/her motivation to prepare for the test. Similarly, Bright and Von Randow (2004) found that low-proficiency students were unwilling or unable to take steps to improve their scores because of time concerns, workload, or the stress of coping with new academic environments, although they accepted the accuracy of their DELNA (Diagnostic English Language Needs Analysis) scores and knew that language proficiency could affect their academic progress. In contrast, Ferman (2004) and Shohamy et al. (1996) found that students with lower abilities, given their belief that cramming improves their scores, engage in more intense learning for the test than their counterparts. The higher-ability students were already eager to learn, even without the push of the test.

What is lacking in the aforementioned empirical studies is whether learners' viewpoints of the tests are related to their language proficiency and therefore lead to a variety of learning strategies among individuals. This study therefore aims to investigate learner washback variability by exploring the relationship between learners' viewpoints of the tests, learning strategies, and test performance.

Research Context

Since 2003, Taiwan's Ministry of Education has encouraged universities and colleges of technology to set English thresholds for graduates to generate a level of English proficiency sufficient to meet the anticipated needs of both domestic and international job markets. In response, the number of four-year technical universities/colleges in Taiwan implementing an English certification exit requirement policy for non-English majors increased from 5% in 2003 to 90% in 2012 (Roever & Pan, 2008; Pan, 2010, 2012). To meet the English certification exit requirements, students must choose from an array of external English proficiency tests such as the TOEIC®, TOEFL®, IETLS™, the General English Proficiency Test (GEPT), and the College Student English Proficiency Test (CSEPT) and reach a certain level or score in order to graduate. However, to assist students who did not pass the required English proficiency tests, "support/alternative/complementary measures" (translated literally from Mandarin) are provided. These measures include taking remedial classes and/or taking an internal in-house test.

The tertiary institute recruited for this study established its English certification exit requirements in 2008. At the time of this study, fourth-year university students were not under the policy, while the remaining three student cohorts had to meet the exit requirements in order to graduate. Students are required to show the school their English certification test results by their fourth year at the latest. To assist their students with passing any of the English certification tests, in addition to using the regular English reading textbooks adopted before the establishment of the exit requirements, students were instructed with either GEPT-related teaching materials or TOEIC-related teaching materials. The school library also offers online websites on which students can practice test-related questions that often appear on various English certification tests.

Research Questions

It is the intent of this study to investigate various learning washback effects on students under different policies (exit vs. non-exit requirement policy), different years of study (1st through 3rd year), and different levels of proficiency. For simplicity, “non-exit students” refer to those who are not under the English certification exit requirements and “exit students” refer to those who are under such requirements.

1. Did exit students outperform non-exit students on the standardized tests?
2. Are exit students and non-exit students motivated to take the English proficiency tests required by the exit policy?
3. Do exit students and non-exit students use different learning activities to study English after class?
4. Are exit students’ after-class learning activities affected by their years of study and levels of proficiency?
5. Are exit students’ viewpoints of the tests affected by different levels of proficiency?

Methods

Rationale

Alderson and Wall (1993, p. 117) point out that washback effects exist when “teachers and learners do things they would not necessarily otherwise do because of the test.” Similarly, Messick (1996, p. 241) describes washback as “the extent to which the introduction and the use of a test influences language and teachers to do things they would not otherwise do that promote or inhibit language learning.” To investigate test effects, Wall and Horak (2007) state the importance of baseline studies in the investigation of test impact in that data collected at the baseline can be used as a point of comparison to assess if any change caused by the test has taken place. This study, therefore, has employed baseline data as a point of comparison to investigate any effects tests have had on students’ motivation, learning activities, and test performance.

Subjects

The technical university recruited for this study has 11 disciplines, 8 of which belong to the business college, the remainder belonging to the science college. In order to make the samples representative, stratified sampling, in which the population was divided into subgroups (strata) and random samples were selected from these strata (Castillo, 2009), was utilized. Based on this probability sampling technique, three disciplines were chosen from the business school and one was chosen from the science school. In total, 587 students were recruited, and a breakdown of these four student cohorts is shown in Table 1.

As mentioned in the section “RESEARCH CONTEXT,” these students attended the same technical university, which established its English certification exit requirements in 2008. At the time of the study, only the fourth-year students had been admitted before the English exit requirement regulation had been instituted. As shown in Table 1, 457 students (referred to as “exit students”) were enrolled under the English certification exit requirements, and 132 students (referred to as “non-exit students”) were enrolled without such a policy.

To explore research questions 4 and 5, exit students were grouped into three proficiency levels. Exit students whose scores are below the sum of the mean scores minus one standard deviation are considered low-proficiency students ($n=68$, 14.9%), those whose scores are above the sum of the mean scores plus one standard deviation are high-proficiency students ($n=77$, 16.8%), and those whose scores are between the low- and high-proficiency students are the intermediate-level proficiency students ($n=312$, 68.3%). The number of exit students in the three proficiency levels follows the curve of standard distribution.

Table 1. Subject profiles

		Exit students ($n=457$) ($n=457$)		Non-exit students ($n=132$)	
Gender	Male	164	35.9%	39	29.5%
	Female	293	64.1%	93	70.5%
Years of study	First	180	39.4%		
	Second	149	32.6%	Not applicable	
	Third	128	28.0%		
	Fourth	Not applicable		132	100.0%
Disciplines	Marketing	119	26.0%	32	24.2%
	Management	110	24.1%	30	22.7%
	Leisure Management	112	24.5%	36	27.3%

		Exit students (n=457) (n=457)		Non-exit students (n=132)	
	Computer Science	116	25.4%	34	25.8%
Test-taking experience (school-made test excluded)	Yes	122	26.7%	46	34.8%
	No	335	73.3%	86	65.2%
Pass certification exit requirements?	Pass GEPT	112	24.5%		
	Pass TOEIC	22	4.8%	Not applicable	
	Pass school-made test	43	9.4%		
	Not yet pass	280	61.3%		

Instruments

Two student questionnaires, one practice GEPT test, and one practice TOEIC test were utilized to explore the five research questions.

The GEPT is a four-skill (listening, reading, speaking, and writing) general English proficiency examination commissioned by Taiwan's Ministry of Education in 1999 to encourage the study of English in a communicative way (Wu & Wu, 2007). The TOEIC, which is comprised of a Listening Section and a Reading Section, "measures the everyday English skills of people working in an international environment" (ETS, 2012). Due to the limited amount of time, the recruited students were required to take the listening and reading sections of the practice GEPT test. In addition, the practice TOEIC test was shortened from 200 to 100 questions, where the first half of the questions in each section were used for this study.

The two student questionnaires were developed based on the results of washback studies conducted by Cheng (2005), Stoneman (2006), Green (2007), and Pan (2010). The intent of the first questionnaire was to determine whether there is a difference between exit and non-exit students in regard to the following two issues.

- Intention to take external English proficiency tests and motivation for taking such tests (in the multiple-choice format)
- Various types of after-class learning activities ranked by students on a 5-point Likert scale (always, usually, often, seldom, never; the Cronbach's α reliability for the 24 items on a 5-point Likert scale was 0.913.)

The intent of the second questionnaire was to determine exit students' viewpoints of the GEPT and TOEIC tests on a 5-point Likert scale (from strongly agree to strongly

disagree) with regard to their importance in enhancing students' language skills and motivation to English study. The Cronbach's α reliability for the 28 items on a 5-point Likert scale was 0.928.

Data Collection and Analysis

Students were asked to fill in the questionnaires translated into Chinese, and take the GEPT test as the first task and the TOEIC test two months later as the second task. The Chinese version of the student questionnaires was delivered by the researcher to student respondents to facilitate the process of completing the questionnaire and put the participants at ease (Johnson & Christensen, 2000). Those who completed both tasks were given reward vouchers.

Before the main study commenced, 10 exit students and 5 non-exit students were asked to complete the questionnaires. One student from each group was subsequently asked if they were confused by any of the wording of the questionnaire items. Both groups of students agreed that each question was clear and comprehensible, so no substantial changes were made. Out of 610, 21 students did not finish the second task, resulting in a completion rate of 96.3%.

Independent exploratory factor analyses using principal component analyses with varimax rotation were conducted on student questionnaires for exit and non-exit students. The factorability of the items in the two student questionnaires was examined by the Kaiser-Meyer-Olkin measure and Bartlett's test of sphericity. Factor loadings lower than .3 were deleted and were not counted toward any factors.

The first questionnaires for both exit and non-exit students

The factor structure for after-class learning activities used by exit and non-exit students in the first questionnaire (24 items, $\alpha=.914$, see appendix 1 for details) was grouped into two major categories: a) types of test-preparation activities and b) types of language skill-building activities. The factor structure of the 1st student questionnaires with Cronbach's α and the number of items was listed in Table 2, as follows:

Types of test-preparation activities

Exit and non-exit students' test-preparation activities were grouped into three types:

- (1) AOTP: autonomous out-of-school test preparation, where students practiced self-purchased 4-skill test-preparation materials.
- (2) ASTP: autonomous school website test-preparation, where students practiced reading and listening tests on the online programs provided by the school
- (3) ITP: in-class assigned test-preparation, where students practiced the mock tests (e.g., listening, reading, speaking and writing) assigned by their instructors.

Types of language skill-building activities

The language skill-building activities engaged in by exit and non-exit students were grouped into three types:

- (1) APR: autonomous practice of receptive skills, where students read magazines, newspapers, novels, browsing English-language websites, listening to English broadcast programs
- (2) APP: autonomous practice of productive skills, where students employed communicative activities to practice writing and speaking skills, such as keeping an English-language diary, talking with foreigners, joining English conversation clubs, and practicing English with teachers
- (3) RPRP: required practice of both receptive and productive skills, where students were asked by their instructors to read textbooks, memorize vocabulary and idioms, study grammar and sentence patterns, listen to audio versions of textbooks, and practice writing short sentences as after-class exercise

Table 2. First questionnaire: Exit and non-exit students' after-class test-preparation and language skill-building activities

After-class activities	Factors	Number of items used	Reliability(α)
Test-related	AOTP Autonomous out-of-school test-preparation	4	.939
	ASTP Autonomous school online test-preparation	2	.892
	ITP In-class assigned test – preparation	4	.894
Language skill-building (non-test related)	APR Autonomous practice on receptive skills	5	.822
	APP Autonomous practice on productive skills	4	.612
	RPRP Required receptive and productive skills practice	5	.805

The second questionnaires for exit students

The factor structure for exit students' viewpoints of the GEPT and TOEIC in the second questionnaire (26 items, $\alpha=.904$, see appendices 2 and 3 for details) was grouped into three categories, as shown in Table 3. The three categories are as follows:

- (1) ELS (enhancement of language skills): Preparing for GEPT and TOEIC can enhance students' four language skills
- (2) EEM (enhancement of extrinsic motivation): Passing GEPT and TOEIC can help the students to graduate, continue their education, enhance their workplace competitiveness, and therefore is important
- (3) EIM (enhancement of intrinsic motivation): Preparing for GEPT and TOEIC can help the students set learning goals and enhance their study time, interest in studying English, and motivation to study English.

Table 3. Second questionnaire: Exit students' views of the GEPT and TOEIC tests

View	Factors	Number of Items used	Reliability(α)
GEPT/TOEIC	ELS Enhancement of language skills	4/4	.455/.756
	EEM Enhancement of extrinsic motivation	4/4	.702/.743
	EIM Enhancement of intrinsic motivation	5/5	.873/.868

Five types of data analyses were adopted to answer the five research questions. First, descriptive analyses of the demographic variables were conducted for both questionnaires. Second, exploratory factor analyses were conducted to plot the major constructs of the two questionnaires. Third, independent sample t-tests were used to explore the relationships within the two groups of students (exit and non-exit students) in regard to their test performance and after-class learning activities. Fourth, one-way ANOVA test procedures were employed to determine the differences across exit students in different years of study and with different levels of language proficiency. Finally, the Chi-square test was employed to analyze the nominal data in regard to exit and non-exit students' intention to take the English proficiency tests. Statistically significant differences with alpha were set at .05 for the data. In addition, effect sizes using Cohen's d for the statistically significant items were calculated to better gauge how the different groups of students varied. Table 4

summarizes the participants involved, the instruments adopted, and the procedures of data analysis in this study.

Table 4. Methodologies

Participants	Instruments	Data Analysis
457 exit students 132 non-exit students	75-item GEPT practice test 100-item TOEIC practice test	Independent sample t-test (interval data) ANOVA test (interval data)
457 exit students 132 non-exit students	1st student questionnaire	Descriptive statistics Factor analysis Independent sample t-test ANOVA test Chi-square (nominal data)
457 exit students	2nd student questionnaire	Descriptive statistics Factor analysis Independent sample t-test ANOVA test

Results

The following sections present the results of the five research questions.

Research Question 1: Did exit students outperform non-exit students on the standardized tests?

As shown in Table 5, the exit students' overall exit test scores were better than those of the non-exit students, and the difference between their GEPT listening scores was statistically significant, $t(587) = 2.056, p < .05$. However, the effect size is small ($Cohen\ d = .263$), the implication of which is that the test-driven policy did not have much effect on the improvement of test scores.

Table 5. GEPT and TOEIC scores of exit and non-exit students

		Exit students (n=457)		Non-exit students (n=132)	
		M	SD	M	SD
GEPT scores	Listening*	86.81	16.89	82.39	16.66
	Reading	81.57	17.39	80.98	17.47

		Exit students (n=457)		Non-exit students (n=132)	
		M	SD	M	SD
	Total	168.38	30.79	164.37	30.64
TOEIC scores	Listening	140.41	46.86	137.12	46.17
	Reading	89.40	48.69	86.02	48.34
	Total	229.60	76.09	223.10	79.91

* $t(587)=2.056, p=.040$

Research Question 2: Are exit students and non-exit students motivated to take the English proficiency tests required by the exit policy?

As shown in Table 6, there was a significant difference between the number of exit and non-exit students who intended to take the external English certification examinations, $\chi^2 (8, N=589) = 153.60, p < .05$. As shown in Table 7, the major difference between exit and non-exit students in their motivation to take the English proficiency tests is that 60% of exit students took the English certification tests in order to meet the exit requirements, while merely 5% of non-exit students had such an intention.

Table 6. Exit and non-exit students who intend to take English certification exams

	Intend to take exams	Do not intend to take exams	Total
Exit students	422 (71.8%)	34 (5.8%)	457 (77.6%)
Non-exit students	60 (10.2%)	72 (12.2%)	132 (22.4%)

Table 7. Exit and non-exit students' motivation to take English certification exams

Motivation to take English certification exams	Exit students (n=457)		Non-exit students (n=132)	
	n	%	n	%
1. To pass exit requirements for graduation	253	60.0	3	5.0
2. To verify one's English proficiency level	183	43.4	34	56.7
3. To enhance one's English proficiency level	271	64.2	49	81.7
4. To enhance one's competitiveness in future education	123	29.1	9	15.0
5. To enhance one's competitiveness in future workplace	319	75.6	51	85.0
6. To satisfy parents' expectation	34	8.1	4	6.7

Research Question 3: Do exit students and non-exit students use different learning activities to study English after class?

Table 8 shows that English certification exit requirements appear to promote students' autonomous learning for test preparation by making use of in-class test-related materials (ITP) $t(587) = 3.816, p < .05$ and school resources (ASTP), $t(587) = 5.028, p < .05$, and these two types of test-preparation were significantly different from those used by their non-exit counterparts. The effect sizes for ITP and ASTP were .37 and .513, representing small to medium, respectively. However, there was no statistically significant difference between these two groups of students in regard to their language-skill building activities. Traditional language skill building activities (RPRP) such as reading textbooks, memorizing vocabulary and idioms, studying grammar and sentence patterns, listening to audio versions of textbooks, and practicing sentence patterns were most frequently used by both exit and non-exit students.

Table 8. Exit and non-exit students' after-class learning activities

		Exit students (n=457)		Non-exit students (n=132)	
		M	SD	M	SD
Autonomous out-of-school test-preparation	AOTP	1.91	.86	2.05	.99
Autonomous school online test-preparation programs	ASTP*	2.27	.93	1.82	.82
In-class assigned test-preparation	ITP*	2.65	.82	2.34	.82
Autonomous practice on receptive skills	APR	2.00	.63	2.05	.63
Autonomous practice of productive skills	APP	1.51	.49	1.46	.50
Required receptive and productive skills practice	RPRP	2.66	.63	2.68	.69

See appendix 1 for details. Test-related activities: AOTP, ASTP, ITP; language skill-building activities: APR, APP, RPRP

ASTP: $t(587) = 5.028, p=.000$; ITP: $t(587) = 3.816, p=.000$

Research Question 4: Are exit students' after-class learning activities affected by their years of study and levels of proficiency?

As shown in Table 9, first-year exit students had a significantly higher frequency of doing school online test-preparation practice than their second- and third-year counterparts. Second-year exit students decreased the frequency of doing test-preparation practice but increased the frequency of doing non-test-related type of practice on receptive skills. The frequency of third-year exit students practicing self-purchased test-preparation materials and receptive skills was markedly higher than that of first-year exit students. The implication of this finding is that as the time to submit their test scores approaches, the higher the frequency of both autonomous practice on test preparation and receptive skills.

Table 9. After-class learning activities from first- to third-year exit students

		1 st -year Exits (n=180)		2 nd -year Exits (n=149)		3 rd -year Exits (n=128)	
		M	SD	M	SD	M	SD
Autonomous out-of-school test-preparation	AOTP*	1.69	.76	1.98	.94	2.14	.83
Autonomous school online test-preparation	ASTP*	2.50	.99	1.93	.79	2.34	.89
Required in-class test-preparation	ITP	2.71	.81	2.61	.87	2.63	.78
Autonomous practice on receptive skills	APR*	1.90	.60	2.07	.65	2.06	.63
Autonomous practice on productive skills	APP	1.49	.43	1.50	.52	1.55	.54
Required receptive and productive skills practice	RPRP	2.71	.67	2.63	.65	2.63	.53

See appendix 1 for details. Test-related activities: AOTP, ASTP, ITP; language skill-building activities: APR, APP, RPR

AOTP: $F(2,455)=11.578, p=.000$; ASTP: $F(2, 455)=16.519, p=.000$;
APR: $F(2,455)=3.578, p=.029$

As shown in Table 10, proficiency levels did not bring about significant differences in the frequency of doing online test-preparation practice on the school website. However, proficiency levels brought about significant differences in the frequency of doing autonomous and required practice on self-purchased test preparation materials and language-skill building activities. In other words, higher-proficiency students utilized a variety of resources for test preparation (self-purchased materials, and in-class test-related materials) and employed various types of practice of receptive and productive skills as compared to low-proficiency exit students.

Table 10. After-class learning activities of exit students with different levels of language proficiency

		High Exits (n=68)		Intermediate Exits (n=312)		Low Exits (n=77)	
		M	SD	M	SD	M	SD
Autonomous out-of-school test-preparation	AOTP*	2.05	.76	1.94	.83	1.66	1.03
Autonomous school test-preparation	ASTP	2.41	1.04	2.22	.91	2.37	.92
In-class test-preparation	ITP*	2.75	.93	2.69	.80	2.41	.77
Autonomous practice on receptive skills	APR*	2.28	.71	1.99	.60	1.78	.58
Autonomous practice on productive skills	APP*	1.73	.60	1.48	.46	1.45	.49
Required receptive and productive skills practice	RPRP*	2.80	.59	2.67	.62	2.50	.67

See appendix 1 for details. Test-related activities: AOTP, ASTP, ITP; language skill-building activities: APR, APP, RPRP

AOTP: $F(2,455) = 4.312, p = .014$; ITP: $F(2,455) = 4.260, p = .015$;
 APR: $F(2,455) = 11.714, p = .000$; APP: $F(2,455) = 7.749, p = .000$;
 RPRP: $F(2,455) = 4.392, p = .013$

Research Question 5: Are exit students' viewpoints of the tests affected by different levels of proficiency?

As shown in Table 11, the exit students had significantly different viewpoints of the English certification exit tests, the GEPT and TOEIC. When asked (1) whether passing the GEPT and the TOEIC can help them to graduate, continue their education, and enhance their workplace competitiveness, and (2) whether preparing for GEPT and TOEIC can help them to set learning goals and enhance their study time, interest in studying English, and motivation to study English, exit students held a more favorable view of the TOEIC than of the GEPT. They assumed that the TOEIC was more beneficial in regard to enhancing their extrinsic motivation (EEM), $t(457) = 9.06, p < .05$, and intrinsic motivation (EIM), $t(457) = 6.15, p < .05$.

Table 11. Exit students' views of the GEPT and TOEIC

		M	SD	t	p-value
Exit students (n=457)	GEPT ELS	3.89	.757	.22	.82
	TOEIC ELS	3.88	.559		
	GEPT EEM	4.00	.603	-9.06	.00*
	TOEIC EEM	4.25	.518		
	GEPT EIM	3.58	.759	-6.15	.00*
	TOEIC EIM	3.71	.730		

See appendices 2 and 3 for details. ELS: enhancement of language skills; EEM: enhancement of extrinsic motivation; EIM: enhancement of intrinsic motivation

As shown in Table 12, the intermediate- and high-proficiency exit students had more favorable views of the tests than low-proficiency students. However, higher-proficiency-level exit students appeared to believe that preparing for these exit tests are more likely to enhance their language skill proficiency and improve their intrinsic motivation, compared to lower-proficiency-level exit students.

Table 12. Views of different levels of exit students

	High Exits (n=68)		Intermediate Exits (n=312)		Low Exits (n=77)	
	M	SD	M	SD	M	SD
*GEPT ELS	3.85	.55	3.95	.82	3.67	.56
*TOEIC ELS	3.94	.52	3.94	.55	3.62	.55
GEPT EEM	3.91	.68	4.02	.58	4.02	.59
TOEIC EEM	4.25	.37	4.27	.52	4.14	.58
*GEPT EIM	3.65	.87	3.64	.71	3.32	.76
*TOEIC EIM	3.85	.71	3.77	.71	3.35	.72

See appendices 2 and 3 for details. ELS: enhancement of language skills; EEM: enhancement of extrinsic motivation; EIM: enhancement of intrinsic motivation (ELS), GEPT $F(2,455) = 4.391, p = .013$; TOEIC $F(2,455) = 10.331, p = .000$; (EIM), GEPT $F(2,455) = 5.717, p = .004$; TOEIC $F(2,455) = 11.765, p = .000$.

Discussion

Test impact on test performance, motivation to take tests, and test-related and language skill-building activities

The major aim of this study was to investigate how exit tests influence learners' motivation, learning activities, and test performance. The English certification exit requirements have brought about several effects. First, the exit students' overall exit test scores were better than the non-exit students,' and the difference between their GEPT listening scores was statistically significant. Second, the exit requirements for English have motivated students to take external English proficiency tests. Third, English certification exit requirements appear to promote students' autonomous learning for test preparation by making use of in-class test-related materials and school resources, and the use of these two types of test-preparation by exit students was significantly different from their use by non-exit counterparts. Fourth, such a test-driven policy did not change students' language skill-building activities to any significant degree because there was no significant difference between exit and non-exit students in regard to the frequency of their non-test-related practice on receptive and productive skills. Since the type of school online test-preparation primarily involved practice of receptive skills such as listening and reading, this may at least partially account for exit students scoring significantly higher on the GEPT listening test than non-exit students. However, increased scores resulting from test-preparation activities may not necessarily represent a corresponding increase in language abilities because such test-taking skills may merely help test-takers who do not possess the target skills to guess the right answers, therefore leading to the increase in their scores (Messick, 1982).

The GEPT and TOEIC, according to their developers, are designed to test learners' language abilities in their daily lives and in the business world respectively, so both receptive and productive skills are assessed. However, there is no significant difference between exit and non-exit students in terms of their frequency of adopting traditional methods and communicatively-oriented methods to study English. These traditional methods such as reading textbooks, memorizing vocabulary and idioms, studying grammar and sentence patterns, listening to audio versions of textbooks, and writing short sentences were actually reported by these two groups of students to be the most frequently used. Communicatively-oriented methods such as keeping an English-language diary, talking with foreigners, joining English conversation clubs, and practicing English with teachers were used the least often by these two groups of students (see Table 8). In view of this, it is a difficult task to change students' study habits by means of a test-driven policy, as found in the studies by Cheng (1998), Stoneman (2006), Chu (2009), and Xie (2013). For example, Xie (2013) found, through the results of 873 student questionnaires, that rehearsing test-taking skills was the most frequently used test preparation practice, but "learning strategies focused on the development of language skills via extensive and functional uses of English language were used scarcely" (p. 210). This result may be partially due to the fact that English, in these researchers' contexts, is

learned as a foreign language, not a second language. English is studied as a “school subject,” and learners do not have many opportunities to utilize what they have prepared for the tests in their daily lives, especially in their school lives. In other words, the English certification exit requirements appear to have promoted “assessment of learning,” but it is a challenging task to determine how to utilize such a test-driven policy to promote “assessment for learning,” where learners can use what they have prepared for the tests in the real world (Cheng et al., 2010).

Test impact on learner washback variability

Another objective of this study was to test Alderson’s and Wall’s hypotheses regarding learner washback variability: 1) A test will have washback for some learners but not for others, 2) A test will influence the rate and sequence of learning, and 3) A test will influence the degree and depth of learning. Based on the findings drawn from exit students’ views on different tests, their years of study, and their levels of English proficiency, this study appeared to confirm these three hypotheses.

Because the TOEIC is an international test (whereas the GEPT is a local test), the exit students probably held a more favorable view of the TOEIC than they did of the GEPT, assuming that the TOEIC was more beneficial in regard to enhancing their extrinsic motivation and intrinsic motivation. According to student questionnaire 2 (see Table 3 and appendices 2 & 3 for details), extrinsic motivation refers to the stakes of the test, where passing the test can help one to graduate, to continue his or her education, or to enhance his or her job readiness, and therefore it is an important test. Intrinsic motivation refers to preparation for the test being able to help one to set his or her learning goals or increase the amount of time one studies or one’s motivation to study English. This finding is consistent with findings from Stone’s study (2006), in which students were more motivated to prepare for the higher-status exam (the International English Language Testing System) than the lower-status test (the Graduating Students Language Proficiency Assessment). However, the students in both studies preferred traditional methods of preparing for the test such as “doing practice exercises using past papers and/or test preparation books” (p. 398). Therefore, for the students participating in these two studies, tests do not influence their strategies for learning English or preparing for the test.

Intermediate- and high-proficiency exit students had more favorable views of the exit tests than their low-proficiency counterparts did. Significant differences were found in exit student views of whether or not these exit tests can improve their language skill proficiency and enhance their intrinsic motivation. It is interesting to note that there were no differences of opinion in regard to whether these exit tests can enhance extrinsic motivation among these three proficiency levels of exit students. In other words, the three proficiency levels of students all contended that these two exit tests were important for their graduation, job applications, and further education. The test they took made no difference to these low-proficiency exit students other than its test stakes, due to the fact that the test requirement was a source of extrinsic motivation that resulted in increased test preparation in the hope of passing the test. This probably explains why lower-proficiency exit students primarily engaged themselves in certain types of test-specific practice such as after-

class learning activities, whereas higher-proficiency exit students adopted various test-related and language skill-building activities (see Table 10). In addition, high- and intermediate-proficiency exit students are also aware of which test can enhance their intrinsic motivation more and which test can enhance their language skills more. This finding may suggest that when teachers and administrators are choosing exit tests for low-proficiency students, the stakes of the test are what these people need to take into account due to the fact that the low-proficiency students will choose learning activities to practice based upon whether they regard the test as important for their ability to graduate, to seek employment, or to further their education.

The types of learning activities adopted by the first-through-third-year exit students appear to confirm Alderson and Wall's hypothesis regarding how a test influences the rate and sequence of learning. Alderson and Wall wanted to determine how quickly and in what order tests would influence learning. First-year exit students conducted school online test-preparation practice to a significantly higher degree. The frequency to which second-year exit students did test-preparation practice decreased, but they spent an increased amount of time on non-test-related practice of receptive skills. Compared to their first-year counterparts, third-year exit students had a significantly increased frequency of doing various types of test-preparation and non-test-related types of practice of receptive skills (see Table 9). These findings suggest that, as time progresses, exit requirements could influence the frequency and order of how students conduct both test-related and non-test-related practice. The results for third-year exit students confirmed the results of Shohamy et al.'s study (1996), in which the number of test-related activities in which participants engaged increased as the date of the test approached. But the results for the second-year exit students showed a decrease in frequency of test preparation. This decreased amount of test preparation frequency for the second-year students is probably because no test-preparation classes were offered for them, and there was still some time before they were required to submit their test scores to meet the exit requirements.

The different proficiency levels of exit students also confirmed Alderson and Wall's hypothesis regarding how a test influences the degree and depth of learning. According to Alderson and Wall, degree and depth refer to the quality and quantity of learning influenced by tests. Although proficiency levels did not indicate significant differences in the frequency of online test-preparation practice, English certification exit requirements appear to affect both quality and quantity of higher-proficiency students' learning activities (such as a variety of resources for both test-related and non-test-related activities). Furthermore, low-proficiency students did not spend as much time on certain types of test preparation and language-skill building activities as intermediate- and high-proficiency students did, but the English certification exit requirements encouraged them to make efforts to utilize school resources for test preparation, as the intermediate- and high-proficiency students did. This could be considered to be a positive test effect for low-proficiency students, because at least they spent some of their time on test preparation.

Taking all of the results from the two practice tests and the two student questionnaires into consideration, there are several pedagogical implications and suggestions.

First, given the insignificant difference in test scores on reading skills between exit and non-exit students, other than test-specific practice, this finding indicates that test preparation did not make much difference in reading skills for the students. Due to the fact that the GEPT was established to encourage the study of English in a communicative way (Wu, 2007) and the TOEIC was designed to measure the everyday English skills of people in the workplace (ETS, 2007), students should be encouraged to engage in a greater number of communicatively-oriented activities to practice their reading skills, such as reading magazines, newspapers, and novels, and browsing English-language websites.

Second, intermediate- and high-proficiency students appear to have benefited from the English certification exit requirements more than low-proficiency students because the latter had a lower frequency of doing both autonomous test-specific and language skill-building activities than the former. In addition, low-proficiency exit students had less favorable viewpoints of the two exit tests than their higher-proficiency counterparts. Considering this result, various measures of exit requirements could be employed for students with different levels of proficiency. It is suggested that, in addition to results from a formal standardized exit test, multiple measures (e.g., formative assessments) that match students' proficiency levels and needs must be developed as part of the criteria for exit requirements. This would be much more likely to foster positive effects (Berry & Lewkowicz, 2000; Chu, 2009; Saif, 1999, 2006; Tsai & Tsou, 2009). The formative assessments could include different types of information such as teacher-developed and portfolio assessments to improve the quality and validity of the decisions made based upon student performance (Berry & Lewkowicz, 2000; Henderson-Montero, Julian & Yen, 2003). By introducing multiple measures as part of exit requirements, students will also be encouraged to develop a broader range of language skills than can be assessed on a one-off language test (Pan & Newfields, 2012).

Third, exit students used school resources more often than non-exit students to study English after class. In light of this finding, a variety of school online educational resources covering practice of the four language skills should be provided from which students could choose so that students will have more opportunities of doing both test-related and communicatively-oriented practice.

Conclusion

Through an analysis of student questionnaires and test scores, this study has provided us with a better understanding of learning washback in regard to how tests influence student performance, learning strategies, and motivation. In addition, it highlights the differential effects of exit requirements for different groups of

learners depending on their years of study, proficiency levels, and perceptions of tests.

This study has three limitations that should serve as suggestions for further research. First, its focus is limited to a policy of mandating EFL proficiency in a Taiwanese tertiary context among non-English majors. Whether or not the patterns in this study will generalize to respondents with different educational backgrounds or fields other than English is an open question. Future studies should explore how exit exam policies appear to impact test stakeholders in different learning environments. Second, this study might have been able to investigate substantial and apparent changes brought about by the tests if it had been conducted a few years after the exit tertiary institute had implemented the policy for a longer period. A longitudinal study would surely offer a clearer picture of the long-term effects of this exit exam policy. Subsequent research should include classroom observation and interview data in order to generate a clearer picture of how washback patterns on learners are presented.

About the Author

Yi-Ching Pan is Associate Professor at the General Studies Center, National Pingtung Institute of Commerce.

References

- Alderson, C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13 (3), 280-297. doi: 10.1177/026553229601300304.
- Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Bailey, K. M. (1999). *Washback in language testing*. Princeton, NJ: Educational Testing Service.
- Berry, V., & Lewkowicz, J. (2000). Exit-tests: Is there an alternative? *Hong Kong Journal of Applied Linguistics*, 5(1), 19-49.
- Biggs, J. B. (1995). Assumptions underlying new approaches to assessment. *Curriculum Forum*, 4(2), 1-22.
- Biggs, J. B. (Ed.). (1996). *Testing: To educate or to select? Education in Hong Kong at the cross-roads*. Hong Kong: Hong Kong Educational Publishing.
- Bright, C., & von Randow, J. (2004, September). *Tracking language test consequences: The student perspective*. Paper presented at the Ninth National Conference on Community Languages and English for Speakers of Other Languages (CLESOL 2004), Christchurch.

Buck, G. (1988). Testing listening comprehension in Japanese university entrance examination. *JALT Journal*, 10, 15-42.

Burrows, C. (2001). Searching for washback: The impact of assessment in the certificate in spoken and written English. In G. Brindley & C. Burrows (Eds.), *Studies in immigrant English language assessment*. (Vol. 2, pp. 95-187). Sydney, Australia: National Center for English Language Teaching and Research.

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

Castillo, J. J. (2009). "Stratified sampling – subjects proportionally selected from subgroups." The Scientific Method, Science, Research and Experiments. Retrieved January 10, 2013 from <http://www.experiment-resources.com/stratified-sampling.html>

Chen, L. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the basic competence test in English*. Unpublished doctoral dissertation, The Ohio State University, Ohio, USA.

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

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24 (3), 279-301.

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

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

Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (2nd ed., Vol. 7, pp. 349-364). New York; London: Springer.

Cheng, L., Andrews, S., & Yu, Y. (2010). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong., *Language Testing*, 28(2), 221-249.

Chu, H. Y. (2009). *Stakes, needs and washback: An investigation of the English benchmark policy for graduation and EFL education at two technological universities in Taiwan*. Unpublished doctoral dissertation, National Taiwan Normal University, Taiwan.

ETS Taiwan Representative. (2012). Taiwan TOEIC statistics report, Retrieved Jan 10, 2014 from http://toEIC.com.tw/toEIC_news_page.jsp?type=4

- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 191-210). Mahwah, New Jersey: Lawrence Erlbaum.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge: Cambridge University Press.
- Henderson-Montero, D., Julian, M. W., & Yen, W. M. (2003). Multiple measures: Alternative design and analysis models. *Educational Measurement: Issues and Practice*, 22(2), 7-12.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading, England.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Humphreys, P., & Mousavi, A. (2010). Exit testing: A whole-of-university approach. *Language Education in Asia*, 1(1), 8-22.
- Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23(2), 99-121.
- Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches*. Boston: Allyn and Bacon.
- Lam, H. P. (1994). *Methodology washback and insider's view. Bringing about Change in Language Education*, 83-99.
- Lewkowicz, J., & Zawadowska-Kittel, E. (2008). Impact of the new school-leaving exam of English in Poland. *Research Notes*, 34, 27-31.
- Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Education Psychologist*, 17, 67-91.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- O'Loughlin, K., & Arkoudis, S. (2009). Investigating IELTS exit score gains in higher education. *IELTS Research Reports*, 10. Canberra, Australia: IELTS Australia Pty Ltd.
- Pan, Y. (2010). *Consequences of test use: Educational and societal effects of English certification exit requirements in Taiwan*. Unpublished doctoral thesis. The University of Melbourne.
- Pan, Y., & Newfields, T. (2012). Tertiary EFL proficiency graduation requirements in Taiwan: A study of washback on learning. *Electronic Journal of Foreign Language Teaching, (e-FLT)*, 9(1), 108-122.

- 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* (pp. 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.
- Roever, C., & Pan, Y. (2008). General English Proficiency Test. *Language Testing*, 25(3), 403-408.
- Saif, S. (1999). *Theoretical and empirical considerations in investigating washback: A study of ESL/EFL learners*. Unpublished doctoral dissertation, University of Victoria, Canada.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1-34.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9, 5-29.
- Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigation the effects of test status on students' test preparation behaviours*. Unpublished doctoral dissertation, The Hong Kong Polytechnic University, Hong Kong, China.
- Tsagari, C. (2006). *Investigating the washback effect of a high-stakes EFL exam in the Greek context: Participants' perceptions, material design and classroom applications*. Unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, United Kingdom.
- Tsai, Y., & Tsou, C. (2009). A standardized English Language Proficiency test as the graduation benchmark: Student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), 319-330.
- Valazza, G. (2008). Impact of TKT on language teachers and schools in Uruguay. *Research Notes*, 34, 21-26.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28, 499-509.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge: Cambridge University Press.
- Wall, D., & Horak, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 1, the baseline study*. Princeton, NJ: Educational Testing Service.
- Wall, D., & Horak, T. (2007). Using baseline studies in the investigation of test impact. *Assessment in Education*, 14(1), 99-116.

- Wall, D., & Horak, T. (2008). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 2, coping with change*. Princeton, New Jersey: Educational Testing Service.
- Watanabe, Y. (1996a). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13 (3), 318-333. doi: 10.1177/026553229601300306.
- Watanabe, Y. (1996b). Investigating washback in Japanese EFL classrooms: Problems of methodology. In G. Wigglesworth & C. Elder (Eds.), *The language testing cycle: From inception to washback* (Vol. 13, pp. 208-239). Applied Linguistics Association of Australia Series.
- Watanabe, Y. (2001). Does the university entrance examination motivate learners? A case study of learner interviews. In A. Murakami, (Ed.), *Akita association of English studies., trans-equator exchanges: A collection of academic papers in Honour of Professor David Ingram* (pp. 100-110). Akita, Japan: Akita University.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19-36). Mahwah, New Jersey: Lawrence Erlbaum.
- Wu, J., & Wu, R. (2007, June). *Using the CEFR in Taiwan: The perspective of a local examination board*. Paper presented at the Fourth European Association for Language Testing and Assessment (EALTA) Conference, Sitges, Spain.
- Wu, J. (2008). Views of Taiwanese students and teachers on English language testing. *Research Notes*, 34, 6-9.
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10, 196-218.

Appendix 1

1st student questionnaire: types of exit and non-exit students' after-class test-preparation and language skill-building activities

Directions: In this section, students were asked how often they did the following activities after class and had to tick one of the given boxes for 1) never, 2) seldom, 3) sometimes, 4) usually, and 5) always.

After-class Activities	Factors	Items
Test-related	AOTP Autonomous out-of-school test-preparation	Practice self-purchased test preparations for reading skills
		Practice self-purchased test preparations for listening skills
		Practice self-purchased test preparations for writing skills
		Practice self-purchased test preparations for oral skills
	ASTP Autonomous school online test-preparation	Practice reading tests on the online programs provided by the school
		Practice listening tests on the online programs provided by the school
	ITP In-class assigned test – preparation	Practice the mock listening tests assigned by the instructor
		Practice the mock reading tests assigned by the instructor
		Practice the mock speaking tests assigned by the instructor
		Practice the mock writing tests assigned by the instructor
Language skill-building (non-test related)	APR Autonomous practice on receptive skills	Read English magazines
		Read English newspapers
		Read English novels
		Browse English-language websites
		Listen to English broadcast programs
	APP Autonomous practice on	Read English magazines Read English newspapers

After-class Activities	Factors	Items
	productive skills	Read English novels
		Browse English-language websites
	RPRP Required receptive and productive skills	Read textbooks required by the instructor
		Memorize vocabulary and idioms required by the instructor
		Study Grammar and sentence patterns required by the instructor
		Listen to audio versions of textbooks required by the instructor
		Practice writing short sentences required by the instructor

Appendix 2

2nd student questionnaire: types of exit students' views of the GEPT test

Directions: In this section, students were asked how much they agreed with the following items and had to tick one of the given boxes for 1) strongly disagree, 2) disagree, 3) not sure, 4) agree, and 5) strongly agree.

Types of tests	Factors	Items
GEPT	ELS Enhancement of language skills	Preparing for the GEPT test can enhance my vocabulary and grammar
		Preparing for the GEPT test can enhance my listening and reading skills
		- Preparing for the GEPT test can enhance my writing and speaking skills - Preparing for the GEPT test can help me to learn the communicative ability
	EEM Enhancement of extrinsic motivation	- Passing the GEPT test can help me to graduate - Passing the GEPT test can help me to continue my education - Passing the GEPT test can help me to enhance my workplace competitiveness - Passing the GEPT test is every essential for me
		Preparing the GEPT test can help to set my learning goals
		Preparing for the GEPT test can help to enhance my study time
		Preparing for the GEPT test can help to enhance my interest in studying English
	EIM Enhancement of intrinsic motivation	- Preparing the GEPT test can help to enhance my motivation to study English - Preparing the GEPT test can help me to be willing to study harder for English

Appendix 3

2nd student questionnaire: types of exit students' views of the TOEIC test

Directions: In this section, students were asked how much they agreed with the following items and had to tick one of the given boxes for 1) strongly disagree, 2) disagree, 3) not sure, 4) agree, and 5) strongly agree.

Types of tests	Factors	Items
TOEIC	ELS Enhancement of language skills	Preparing for the TOEIC test can enhance my vocabulary and grammar
		Preparing for the TOEIC test can enhance my listening and reading skills
		- Preparing for the TOEIC test can enhance my writing and speaking skills - Preparing for the TOEIC test can help me to learn the communicative ability
	EEM Enhancement of extrinsic motivation	- Passing the TOEIC test can help me to graduate - Passing the TOEIC test can help me to continue my education - Passing the TOEIC test can help me to enhance my workplace competitiveness - Passing the TOEIC test is every essential for me
	EIM Enhancement of intrinsic motivation	- Preparing the TOEIC test can help to set my learning goals
		- Preparing for the TOEIC test can help to enhance my study time
		- Preparing for the TOEIC test can help to enhance my interest in studying English
		- Preparing the TOEIC test can help to enhance my motivation to study English - Preparing the TOEIC test can help me to be willing to study harder for English

Copyright © 1994 - 2014 TESL-EJ, ISSN 1072-4303
Copyright rests with the authors.