



Tertiary EFL Proficiency Graduation Requirements in Taiwan: A Study of Washback on Learning

Yi-Ching Pan

(yichingpan@yahoo.com.tw)

National Pingtung Institute of Commerce, Taiwan

Tim Newfields

(newfields55@yahoo.com)

Toyo University, Japan

Abstract

With the reputed goal of enhancing students' English proficiency and competitiveness in the global market, an increasing number of universities/colleges in Asia are establishing English proficiency graduation requirements. This study explores how such requirements have impacted 17 tertiary educational institutions in Taiwan. Extensive questionnaire and interview data from students at 8 schools with such requirements and 9 schools without them suggest that mandated EFL proficiency tests have had minimal washback on students. The article concludes by interpreting these results in the light of recent studies on learner washback and suggesting avenues for further research.

1 Introduction

In places such as China (Qi, 2004), Hong Kong (Cheng, 2005), Korea (Choi, 2008), Japan (Watanabe, 2004b), and Taiwan (Chen, 2002), tests are thought to be a major determinant of course designs and classroom practices. Policy makers or school administrators generally believe that tests have much power and therefore quite often try to use them to manipulate or implement educational policies (Shohamy, 2001a, 2001b). With an avowed goal of bolstering students' English aptitude and global market competitiveness, tertiary-level English as a Foreign Language (EFL) graduation requirements are being adopted in many Asian countries. For example, 41 universities in Korea, Japan, Thailand, and elsewhere in Asia (IIBC, 2005, p. 7) have set minimum Test of English for International Communication (TOEIC) score requirements for graduates. Likewise, in 2003, Taiwan's Ministry of Education announced a new policy to encourage all higher education institutions to adopt English proficiency graduation requirements.

Comparatively little research has been conducted regarding the effects of tests on the learning processes, in contrast to the significant number of studies on the effects of tests on teaching. Several researchers (Cheng 2008; Spratt, 2005; Wall, 2000; Watanabe, 2004a) have advocated the need to explore the impact of tests on learners, since they are directly affected by them. This study aims to explore the test effects brought about by such graduation requirements in the Taiwanese tertiary educational framework from the perspectives of students.

The following questions are addressed:

1. How do English proficiency graduation requirements tend to affect students' motivation to study English in class?
2. Do English proficiency graduation requirements seem to influence how students study English outside of class?

1.1 English certification graduation requirements in Taiwan

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.

The Ministry established a list of recommended tests to fulfill this requirement as in Table 1, which are thought to reflect either the CEFR B1 or A2 levels. The list includes the TOEIC[®], TOEFL[®], IELTS[™] and two local tests: the General English Proficiency Test (GEPT) and the College Student English Proficiency Test (CSEPT). The GEPT is a 5-level, four-skill general English proficiency examination commissioned by Taiwan's Ministry of Education in 1999. The CSEPT is 2-level, listening-reading-grammar test for university-level students in Taiwan (Pan, 2009).

CEFR Level	TOEFL		TOEIC	IELTS	GEPT	CSEPT
	Paper-and pencil	CBT				
A2 Waystage	390+ – 456	90–136	350–549	Band 3	Elementary	Level 1 170–229
B1 Threshold	457+ – 526	137–196	550–749	Band 4	Intermediate	Level 2 240+

Table 1: A list of recommended tests to fulfill the English proficiency requirements for tertiary graduates in Taiwan

By 2008, nearly one-third of the universities and colleges of technology in Taiwan have adopted the GEPT as one way to fulfill the English proficiency graduation exam requirements for non-English majors (Roever & Pan, 2008). The GEPT was created at least partially as a response to the belief among educators and employers that Taiwanese English learners possessed a general inability to communicate in English, due to an outdated approach to teaching English that placed far too much importance on grammatical accuracy (Pan, 2009; Wu & Wu, 2007).

2 Literature review

2.1 The concept of washback

The term “washback” is prevalent in language teaching and testing literature as well as general education. Alderson and Wall (1993) define it as “the way that tests are ... perceived to influence classroom practices, and syllabus and curriculum planning” (p. 117). The influence of washback is not limited to students – it compels teachers, students, school administrators, and other test stakeholders “... to do things they would not necessarily otherwise do” (Alderson & Wall, p. 117). Somewhat vaguely, Buck (1988) defines this term simply as the influence that tests exert on teaching and learning. Bachman and Palmer (1996) consider washback to be a feature of a wider process known as *test impact*. They further mention how test impact has both micro effects on specific classrooms and macro effects on societies at large.

For the purpose of this paper, washback is understood to be the effects that tests have on students in terms of the methods they use and motivation to study a second language (L2). These effects are likely to be viewed as positive, negative, or inconsequential by different stakeholders. If curricular goals align well with the material instructors actually teach (the so-called *teaching syl-*

labus) and what students actually wish to learn (the goal of a *learner-led syllabus* [Breen, 1984]) as well as what is tested (the *test syllabus*), then *curricular alignment* is said to take place, and it is hypothesized that under such conditions, washback will tend to be strong. If, on the other hand, test content does not match well with these components, then washback is apt to be either ineffectual or negative.

2.2 Washback effects on instructional approaches

So far, the vast majority of foreign language washback studies have focused on *teacher* perspectives concerning tests. Two contradictory claims have been made: (1) Tests promote instruction of test-oriented activities; and (2) tests promote instruction of communicatively-oriented activities. Each of these claims is briefly discussed below.

2.2.1 Tests promote instruction of test-oriented activities.

The majority of washback studies show that teaching to the test is a common practice. Cheng (2005), Green (2007), Hayes and Read (2003, 2004), and Wall and Horak (2006) found more test-related activities (e.g. offering test-taking tips, doing question analysis) and instruction of test-taking strategies (e.g. formulaic approach to teaching writing, adoption of preparation textbooks, and less attention paid to oral skills) in IELTS/TOEFL preparation classes than in regular classes. Qi's study (2004, 2005, 2007) reveals "cramming for testing" in teaching practices to assist students in doing well on a gate-keeping test, the NMET (National Maturation English Test in China). 57% of the class time was spent on test drilling, 35% on grammar and vocabulary practice, and only 7% on communicative language use.

2.2.2 Tests promote instruction of communicatively-oriented activities.

Lam (1994) investigated whether the Use of English exam in Hong Kong could influence teaching practices. He found that less experienced teachers tended to use different activities to teach oral language, such as creating authentic materials from the mass media, producing meaningful learning activities, encouraging student participation, and adopting an integrated approach to teaching. Hawkey (2006) explored the impact on the Progetto Lingue 2000 (Year 2000 Languages Project) in Italy. The PL2000 aimed to provide foreign language courses to meet the communication needs of students as defined by the Council of Europe's (2001) Framework of Reference for Languages. Students were also encouraged to seek certification of their progress through Cambridge exams like KET, PET, and FCE. He discovered that this curricular alignment program encouraged teachers to use a variety of sources such as "cut-out photographs, self-designed spider games, information-gap hand-outs, audio-cassettes, (and) wall charts" (Hawkey, 2006, p. 143) to improve students' communication skills.

Burrows (2004) and Watanabe (2004b) claim that due to psychological factors (e.g. teaching beliefs, experience, and educational backgrounds), some teachers resorted to explicitly traditional grammar-translation approaches to help students to do well on the tests, while others incorporated communicatively-oriented approaches to develop students' real-life language ability. "Teachers are the final arbiters of policy implementation" (Menken, 2008, p. 401) and it is therefore highly likely that a teacher's perceptions of tests determine the motivation and effort he/she makes to improve pedagogy and impact students' learning behavior (Clark & Peterson, 1986; Lane, Parke, & Stone, 1998). An exploration of the effects of washback on teachers may offer insights about how tests influence learners.

2.3 *Washback effects on learning activities*

Studies on learner washback reveal varied and sometimes contradictory findings. The following section discusses the washback effects on learners' motivation, study time, and learning activities.

2.3.1 *Washback effects on motivation and study time*

Shohamy, Donitsa-Schmidt, and Ferman (1996) contend that a high-stakes test (the EFL exam in their study) promoted learning; however, students' motivation to master the test material significantly decreased after the administration of the test. In contrast, a low-stakes test (an Arabic exam in their study) may have encouraged lower-proficiency students to a greater degree than those students with higher proficiency, because the latter were already eager to learn, even without the push of the test. Similarly, Ferman's (2004) study found that the lower the students' ability level, the more time and effort they allotted to studying for the EFL National Matriculation Test in Israel. Average and lower-ability students hired tutors significantly more often than higher-ability students to help them to prepare for the test. Moreover, Stoneman (2006) investigated how 655 students in Hong Kong prepared for the university graduation exams. Her study revealed that over half of the students were more motivated to prepare for a high-status international exam, but not for a local exam with less status.

On the other hand, Gan, Humphreys, and Hamp-Lyons (2004) found that successful students appeared to be more willing to study for and take an English proficiency test, because they believed it could evaluate their English proficiency. By contrast, unsuccessful students tended to have negative attitudes toward taking the test and experience significant stress and panic. These unsuccessful students tended to feel either unconcerned or resentful toward the test. In the same vein, according to Watanabe (2001) and Chu (2009), students' perceptions of test difficulty or easiness partially determine their levels of motivation and the amount of effort they devoted to preparing for the test. In their studies, low-achieving students tended to be more worried about the test or test requirements than high-achieving students, and these low-achieving students did not prepare for the test until the last minute, if they even did so at all.

Bright and Von Randow (2004) studied how a test of English academic skills affected 18 foreign students in New Zealand. Despite knowing that their test scores would decisively impact their academic careers, low-proficiency students tended to make no effort to improve their scores because of time concerns, workload, or the stress of coping with new academic environments.

Shih (2007) has explored the effects of a graduation exam on 29 Taiwanese university students. Although 58% ($n=17$) of the participants considered that test important, few actually spent time preparing for it. This echoes Bright and Von Randow's point: merely knowing that a test is important is not enough to guarantee that most people will actually study harder for it. Shih speculated that extrinsic factors (e.g. personal factors such as students' part-time jobs), intrinsic factors (e.g. students' learning attitudes), and test factors (the immediate importance of the test; the way the test-driven policy was implemented) were responsible for the phenomena perceived.

These studies imply that tests can serve as a facilitator to motivate students or a detriment to impede their interest in learning. In addition, test stakes (low- or high-stakes), the status of the test, and learners' factors (e.g. level of proficiency and learning attitudes) all play a role in determining the changes students make for the test.

2.3.2 *Washback effects on learning activities*

Stoneman (2006) investigated the methods 655 students in one Hong Kong university used to prepare for two university graduation exams, one local and one international. She found that their methods were similar, and more than half of the students said that they preferred traditional methods such as going through previous exams and/or relying on test prep books.

Similarly, Jou (2010) investigated how 127 students in one Taiwan university prepared for the TOEIC, an exam that they could choose to meet their school graduation requirements for English. Three-fourths of them reported that they adopted traditional test-preparation methods such as practicing test-related questions in either hard- or soft-copy test preparation materials, and taking test-preparation lessons either in or outside of school. Merely 6.3% utilized authentic materials such as watching CNN or listening to the ICRT to prepare for the TOEIC.

Chu (2009) asked 667 students from two Taiwanese universities about their perceived impact of the GEPT graduation exam on their out-of-school practice. The top three learning activities according to a 5-point scale used to prepare for the GEPT test were “listening to English songs” ($M=2.33$), “memorizing vocabulary items” ($M=1.57$) and “watching English TV programs” ($M=1.53$). It is interesting that the students preferred to use entertaining audio-visual materials for test-preparation. However, as indicated by the author, “the low mean scores in the parentheses indicate that students did not do these tasks often and regularly” (p. 122). In other words, the GEPT graduation requirements had a fairly low degree of washback on learning behavior.

Although these learner washback studies could provide some perspective into the effects of tests on learning processes, the participants in most studies were recruited from one or two schools. Such case studies are often criticized for their lack of generalizability (Gall, Gall, & Borg, 2003). In addition, most of these learner washback studies did not adopt an experimental or quasi-experimental research design with control and experimental groups. Using two or more groups will allow us to determine if any differences exist between a situation where there is an exam, and a situation where there is not (Alderson & Wall, 1993; Messick, 1996).

This study seeks to fill a gap by providing a large-scale study with both experimental and control groups.

3 Methodology

This study consisted of two phases: a survey phase, and a structured interview phase. A paper-and-pencil survey was selected because it is the most efficient way to obtain a large sample size relatively quickly (Wall, 2005). Structured interviews made it possible to probe responses in detail. A structured interview was chosen to maximize consistency across respondents (Cohen, Manion, & Morrison, 2007; Watanabe, 2004a).

Two types of comparative approaches are usually conducted to investigate test effects. One approach compares baseline data with data collected in a follow-up study to determine whether tests may have caused the changes (Wall & Horak, 2006, 2007, 2008). The other approach compares test-preparation and regular classes to determine whether there are differences in regard to teaching practices and student learning. If differences exist, they can be taken as evidence of the existence of washback effects (Alderson & Hamp-Lyons, 1996; Green, 2007; Hayes & Read, 2003, 2004).

The present study was conducted after the inception of English certification graduation requirements in Taiwan; so a comparison of the baseline and a follow-up study to determine the consequences brought about by the tests was not viable. Therefore, a comparison of the differences between the schools with graduation requirements and those without graduation requirements will be used to reveal test effects.

The participants, instruments, and procedures for each phase are briefly described.

3.1 Participants

There are 78 four-year technical universities/colleges. A stratified convenience sample of 1415 students was used for the survey phase of this study. This consisted of 737 students from 8 tertiary institutions in Taiwan with English proficiency exam graduation requirements (herein referred to as EGR schools) and 678 students from 9 schools without such requirements (herein referred to as non-EGR schools). A slightly higher percentage (around +3%) of the students at EGR schools has

passed particular English proficiency tests than their counterparts, although the number who passed these tests at both groups of schools was only about one third. The low pass rate suggests the low English proficiency of the students in both groups.

A demographic profile of these two groups appears in Table 2.

		EGR Students <i>n</i>=737	Non-EGR students <i>n</i>=678
Gender	Male	43.8%	59.0%
	Female	56.2%	41.0%
Major	Business/Management	28.0%	67.0%
	Engineering	52.2%	6.8%
	Humanities and Social Science	11.3%	6.9%
	Agriculture	3.0%	0.0%
	Others	5.6%	19.3%
Location of the school	Southern Taiwan	76.9%	80.5%
	Central Taiwan	8.4%	8.7%
	Northern Taiwan	14.7%	10.8%
Pass Exams?	Yes	(151) 38.2 %	(96) 35.3%
	No	(244) 61.8%	(176) 64.7%

Table 2: Student survey informant profiles

For the interview phase of this study, a convenience sample of 71 respondents from the survey sample who indicated a willingness to be interviewed was employed. 38 respondents came from 8 different EGR schools and 33 came from 6 non-EGR schools. A demographic profile of those individuals appears in Table 3.

		Respondents (<i>n</i>)	Percentage (%)
Gender	Male	25	35
	Female	46	65
Major	Business/Management	37	52
	Engineering	19	27
	Agriculture	6	8
	Others	9	13
School Type	EGR schools	38	54
	Non-EGR schools	33	46
Passed Exams?	Yes	14 (10 at EGR schools)	20
	No	57	80

Table 3: Student interview respondent profiles

3.2 Instruments

The two instruments used in this study will be briefly described.

First, a 26-item paper-and-pencil survey was used to determine whether the test requirement has brought a difference between students at the two groups of schools in terms of their motivation for studying English, study time, and out-of-class learning activities (language-skill building/communicatively-oriented activities and test preparation activities). This survey was developed from previously published washback studies (e.g. Cheng, 2005; Shohamy et al., 1996; Stoneman, 2006). The survey contained two types of questions: multiple-choice questions with categorical responses and 5-point Likert scale questions with pseudo-ordinal responses.

Second, the main purpose of the structured interviews with students was to clarify various points from the survey. In addition, the interviews provided participants with an additional opportunity to express their opinions regarding the consequences of the graduation requirements.

3.3 Procedures & data analysis

The procedures and data analysis for each phase of this study are briefly outlined.

Through the main researcher's network with the teachers she knew, the original Chinese version of the translated survey was distributed to 800 students at 8 EGR schools and 750 students at 9 non-EGR schools all over Taiwan. Of the 1,550 surveys distributed, 1,415 valid questionnaires were returned.

SPSS 12.0 was used to analyze the questionnaire data. The Mann-Whitney U test was utilized to check for statistically significant differences with alpha set at .05 for the ordinal data, and the chi-square test for the nominal data. In addition, effect sizes using Cohen's d for the statistically significant items were calculated to better gauge how the two groups of students varied. This significance test should also facilitate any subsequent meta-analyses. The Cronbach Alpha reliability for the 26 items on a 5-point Likert scale was 0.87.

Participants who expressed a willingness to be interviewed in their survey responses were recruited and given a gift in appreciation of their participation. They could choose three different interview formats according to their convenience: telephone, focus groups (4 focus groups with 3-4 students each), or face-to-face individual interviews. Interviews were conducted in Mandarin for optimal communication and to avoid confusion. Each interview lasted about 10-30 minutes. Several follow-up phone interviews took place, if there were incomplete or confusing responses.

All interviews were digitally recorded, transcribed, and then analyzed according to the five analytical-strategy steps proposed by Schmidt (2004). First, "intensive and repeated reading" (p. 254) of the transcribed interviews was done. Second, analytical main and sub-categories were constructed as a fundamental guide. Third, with the assistance of the qualitative software package NVivo 8, all the transcribed data were sorted according to the analytical categories, with the goal of providing examples for the research questions. Fourth, the results were quantified where possible to obtain a preliminary overview of the data. Finally, detailed explanations were presented to draw inferences for the research questions.

4 Results

Each research question shall now be addressed.

4.1 Q1: How do English proficiency graduation requirements tend to affect students' motivation to study English in class?

The survey data in Table 4 suggest that students from both types of schools varied little in terms of motivation for English study. Only 3 of the 12 survey items had statistically significant differences ($p < .05$). One tenth more students from EGR schools agreed that they study English in order to earn certificates ($M=3.90$), to pass the test to graduate ($M=3.84$), or to improve their English for further education ($M=3.63$) than their counterparts. However, the effect sizes for these three items were small ($d < 0.3$), indicating that students at schools with graduation exams tended to be only marginally more motivated to study by English than their peers.

Students' motivation for studying English	Students at EGR schools						Students at non-EGR schools					
	<i>N</i>	<i>M</i>	<i>SD</i>	A (%)	U (%)	D (%)	<i>N</i>	<i>M</i>	<i>SD</i>	A (%)	U (%)	D (%)
(1) To have better future opportunities	735	4.20	0.73	86.7	11.3	2.0	676	4.24	0.70	88.5	10.1	1.5
(2) To learn daily English	735	4.15	0.70	88.2	9.4	2.4	677	4.19	0.70	90.0	7.5	2.5
(3) To communicate with foreigners when traveling overseas	737	4.12	0.79	82.2	14.2	3.5	677	4.12	0.81	82.2	14.5	3.4
(4) For job hunting	732	4.11	0.72	86.6	10.5	2.8	676	4.15	0.71	88.3	8.9	2.8
(5) ⊙*To prepare for proficiency tests and earn certificates	736	3.90	0.82	78.0	14.9	7.1	675	3.67	0.87	65.4	25.2	9.3
(6) ⊙*To pass the test to graduate	731	3.84	0.95	78.2	8.6	13.2	671	3.59	0.96	67.0	15.4	17.6
(7) To be able to watch English movies and listen to English programs	736	3.83	0.91	70.1	21.1	8.8	677	3.85	0.89	71.6	20.4	7.9
(8) To understand professional knowledge written in English	735	3.75	0.85	67.0	25.0	7.9	676	3.72	0.87	65.2	25.9	8.8
(9) ⊙*To enhance English proficiency to continue on to higher education	734	3.63	0.88	61.6	26.8	11.6	677	3.52	0.89	56.0	31.5	12.6
(10) To fulfill parents' expectations	734	3.43	1.00	50.4	31.7	17.8	675	3.43	1.00	52.5	30.4	17.2

A= agree, U= uncertain, D= disagree, ⊙: effect size ≤ 0.3

*: Mann-Whitney U-test, statistically significant difference between the two groups of schools at $p < .05$

Table 4: A comparison of students' motivation for studying English

The differences in Table 4, though small, may be attributed to the influence of graduation requirements. The student interview data below detail how this occurred.

When asked if they had experienced any positive effects as a result of the graduation requirements, 14 of the 24 EGR interviewees indicated the policy motivated them to study "somewhat harder." 10 claimed that they were passive learners who needed external inducements such as the test requirements to compel them to study. Informant S174 commented: "Since I will not be allowed to graduate unless I pass the test, the policy has pressured me into studying English." Although these student informants admitted the graduation requirement policy had exerted some pressure, it brought "more encouragement than pressure," according to informant S63. Six EGR informants, however, expressed anxiety about graduating and candidly acknowledged they had no desire to study English because of the graduation requirements, despite knowing their English levels were low. One informant (S122) also voiced concern that those who could not pass the proficiency test would be looked down upon by their peers. Hence the data suggest the exam pressure enhanced the motivation for some, but impeded others.

On the other hand, only two of the 10 non-EGR interviewees stated that they were motivated to study English in class because they either "need to serve foreign clients when they purchase medicine" or "want to have more job opportunities in the future." The rest expressed that they had "little interest" in English because of their low English proficiency. Some also said they wanted to learn English, but do not really like studying. A few frankly confessed that they were feeling "lazy at the moment."

From the interview findings, the test requirements appear to motivate or push some EGR students to study English to some extent, but this appears to be less so with low proficiency students.

For non-EGR students, their goal of learning has a role to play in how much effort they intend to make to enhance their English.

4.2 Q2: Do such graduation requirements appear to influence how students study outside of class?

The survey data in Table 5 shows how much time students from both groups reported studying English per week outside of class when school was in session.

Around half of the EGR respondents reported spending one hour or more on English each week, compared to merely one third of their counterparts. Although there was a statistically significant difference between these two groups, the effect size was small ($d = 0.2$). This suggests that the difference between both groups regarding time investment in English study was minimal.

	Only before tests	Under 1hr. per week	1-3 hrs. per week	Over 3 hrs. per week	Total Sample
Students at EGR schools	161 (22%)	215 (29%)	292 (40%)	65 (8%)	733 (52%)
Students at non-EGR schools	221 (33%)	224 (33%)	190 (28%)	39 (5%)	674 (48%)
Total Students	382 (27%)	439 (31%)	482 (34%)	104 (7%)	1407 (100%)

$$X^2 = 35.492; df=3, p=.000, d = 0.2$$

Table 5: A Comparison of students' English study time outside of class

Moreover, the survey data in Table 6 compares how the two groups of students reported studying English outside of class. More than half of the students in both groups reported using traditional methods such as text reading, rote memorization, and practicing grammar exercises. Less attention was paid to learning productive skills such as speaking and writing.

Although there were ten statistically significant differences between the two groups, only one had a moderate effect size ($d=0.6$). 20% more students at EGR schools claimed that they “practiced online test-related questions provided by their schools” more frequently than their counterparts. Around 2%–11% more EGR students claimed they frequently or sometimes used the other methods (learning grammar, reading magazines, reading online information, listening to radio programs, joining English clubs, practicing orally with teachers, chatting online in English, and writing emails in English) than their counterparts. In other words, in addition to the cram-oriented traditional methods, more EGR students adopted communicatively-oriented methods to practice each of the four language skills. However, these methods received small effect sizes ($d < 0.3$).

The interviews also suggest that EGR informants were more aware of CALL resources than non-EGR informants. One third ($n=13$) of the 38 interview respondents at EGR schools said that they used their school's language labs or websites for English study as requested by their teachers. In contrast, only 4 of their 33 counterparts mentioned doing this. It is understandable that students at non-EGR schools do not have to pass certification tests to graduate; so, as informant S161 said, “I don't pay much attention to these [educational resources for test-preparation] ... Since I have a heavy load of schoolwork, I don't have the time to use it [on-line test practice on the school website].” Considering this, being better informed of a school's educational resources for test-preparation could be considered a positive effect of the graduation requirements.

Out-of-class English learning activities	EGR schools						Non-EGR schools					
	<i>N</i>	<i>M</i>	<i>SD</i>	F (%)	S (%)	R (%)	<i>N</i>	<i>M</i>	<i>SD</i>	F (%)	S (%)	R (%)
(11) Learning vocabulary	736	3.34	0.92	44.6	37.0	18.5	674	3.30	0.94	41.1	38.0	20.9
(12) *Taking notes	736	3.24	0.98	42.6	33.4	23.9	676	3.36	0.97	46.7	32.8	20.4
(13) Reading textbooks	735	3.11	0.91	33.3	40.8	25.8	675	3.11	0.97	34.4	37.3	28.3
(14) *Learning grammar	729	2.82	0.92	22.7	37.9	39.5	675	2.73	0.91	19.5	35.1	45.4
(15) *Reading online information	734	2.56	1.02	19.1	28.1	52.9	677	2.33	1.00	13.2	23.0	63.8
(16) *Reading magazines	734	2.47	0.91	13.7	31.2	55.2	676	2.29	0.92	10.0	23.4	66.6
(17) Watching movies, and TV programs	735	3.20	1.06	40.5	31.4	28.0	673	3.12	1.06	36.7	35.2	28.0
(18) * Listening to audio versions of the text or radio programs	734	2.56	1.02	21.8	32.7	45.5	677	2.33	1.00	16.8	30.6	52.5
(19) Practicing orally with classmates	736	2.04	0.82	4.8	19.5	75.6	675	1.97	0.80	4.2	16.3	79.6
(20) *Practicing orally with teachers	737	2.02	0.80	4.4	19.0	76.5	675	1.92	0.81	4.8	13.5	81.7
(21) *Chatting on line in English	734	1.84	0.88	5.1	14.4	80.5	677	1.74	0.85	3.9	12.4	83.6
(22) *Joining English clubs	736	1.66	0.80	3.1	9.9	87.0	674	1.59	0.82	4.3	6.5	89.1
(23) Practicing writing short essays or keeping diaries in English	733	1.71	0.81	3.0	11.9	85.1	676	1.65	0.80	3.1	9.8	87.1
(24) *Writing emails in English	735	1.61	0.82	3.5	8.7	87.8	676	1.52	0.76	2.4	7.2	90.4
(25) Going to language cram schools	734	2.35	1.07	16.0	22.9	61.1	677	2.27	1.00	11.9	23.9	64.2
(26) ⊙*Doing practice tests on the school website	731	2.23	0.92	7.8	29.1	63.0	675	1.70	0.86	3.9	11.7	84.5

F = frequently, S= sometimes, R=rarely;

⊙: moderate effect size ($d = 0.6$), all other effect sizes were small $d \leq 0.3$

*: Mann-Whitney U-test, statistically significant difference between the two groups of schools at $p < .05$

Table 6: A comparison of students' methods of studying English outside of class

5 Discussion

The above differences suggest four slight changes brought about by the English graduation requirements:

- 1) increased motivation for English study,
- 2) more time allocated to English study,
- 3) more variation in the methods adopted to study English, and
- 4) more test-related practice engaged.

5.1 Test effects on motivation and study time

Gardner and MacIntyre (1993) suggest that motivation consists of three components: (1) a desire to achieve a goal, (2) an effort towards that direction, and (3) a feeling of fulfillment when a task is completed (p. 2). Though mandating English proficiency exams may have enhanced the first two components (e.g. EGR students' higher desire to pass the proficiency tests and more al-

lotted time to English study), there is no evidence suggesting that any sense of task fulfillment was increased, considering the low pass rate of students at EGR schools (see Table 2). Thus, the effect of the English certification graduation requirements on student motivation appears to be partial at best.

In addition, Harlen and Crick (2003) remind us that motivation is complex and associated with concepts such as self-esteem, self-efficacy, effort, and self-regulation (p. 169), which is not something a mere test can engender. Indeed, this might explain why gains in EGR students' avowed motivation to study were nominal, and why changes in the amount of time they reputedly studied were also so modest. The implication is that, to a large extent, English certification graduation requirements have not pushed students to study hard.

Moreover, as reported in the learner washback studies, some EGR students' motivation to study English may be impeded by their low English proficiency. The finding from the EGR student interview data tends to correspond with Fransson's (1984) contention that learner motivation does not always have a linear relationship with learner behavior. Fransson asserts that learners' performance declines when they reach the point at which the challenge of the task negatively affects their motivation. Those who find the test difficult are prone to become frustrated and discouraged and end up unprepared.

Although EGR students spent only marginally more time on English study than non-EGR students, this could be considered as one of the positive effects brought about by the graduation requirements. Given the fact that students are not considered to be active learners (according to the low mean scores for the out-of-class learning activities in Table 6), even this minor increase in the amount of time students spent studying seems encouraging.

5.2 Test effects on learning activities

This study makes it clear that the test requirements did not lead to a noteworthy amount of "studying for the test," a phenomenon often reported in examination-oriented societies (Chern, 2002; Lai, 2003; Tsai & Tsou, 2009). However, the graduation requirements may have encouraged a small amount of learning through communicatively-oriented and test-preparation approaches.

In addition to reading skills, which have been emphasized in English classes, EGR students tended to adopt various language learning activities to practice their listening, speaking and writing skills more than non-EGR students. As a consequence, there was a slight increase in their foreign language productive skills. However, there was no dramatic change in students' learning activities, because for most of the respondents in this study the old habits of traditional, non-communicative study seemed to remain entrenched. This finding is consistent with Stoneman's (2006) conclusion that tests do not influence students' strategies for learning English or test preparation. As mentioned earlier, learner washback is mediated in part by teachers (Burrow, 2004; Watanabe, 2004b). As a result, student preferences for traditional methods of reading texts and memorizing grammatical rules, vocabulary, and phrases can likely be attributed to a preference among Taiwanese teachers for teacher-centered instruction and text explanations (Cheng, 2006). Moreover, as found in the student interview data, one third of EGR students utilized their schools' CALL resources for test preparation, as requested by their teachers. Partially because of teachers' belief that helping students earn English certificates would benefit them in the future, these teachers assigned their students out-of-class test-preparation practice (Pan, 2010). At any rate, the graduation requirements did seem to increase the amount of test-related practice by EGR students to a certain extent.

6 Conclusions

This study concludes by highlighting the implications of this research for EFL teaching and learning, acknowledging some of its limitations, and suggesting avenues for further research.

6.1 Implications for EFL teaching and learning

A sobering finding is that simply relying on external tests alone to enhance motivation appears to be ineffective. Tests must complement other measures, such as collaborative efforts by teachers, who face students directly and therefore have a greater impact on them by their selection of curricula (Menken, 2008). Investigation of students' needs for their further studies or for future employment could also enhance their interest and increase their motivation. According to Saif (1999, 2006), test effects on learning would be strong and positive, if a language test could match learners' language needs. In addition, the passing scores/levels should be neither so lax that most students do not even bother to study, nor so rigorous that they are prone to become frustrated and discouraged and end up unprepared. Moreover, the alignment of curricula with test content may be one possibility teachers can consider so that they could focus both on receptive and productive skills in class. Finally, this study confirms the contention of other washback studies that standardized tests are not a panacea that will always succeed in changing students' study habits. In order to encourage students to participate in extracurricular activities (e.g. writing emails, reading English magazines, joining English clubs, online chatting, and listening to English radio programs) to increase their opportunities to practice a foreign language, assessment should include portfolio of work reflecting their ability to use English in a variety of contexts (Berry & Lewkowicz, 2000; Chu, 2009; Hsu, 2010; Tsai & Tsou, 2009). This could be developed as part of their graduation criteria. By introducing multiple measures to fulfill their graduation requirements, students will also be encouraged to develop a broader range of language skills than can be assessed on a one-off language test.

6.2 Limitations of this study and further avenues to explore

This study has three limitations that indicate directions for further research. First, this study was conducted at a period when many institutions in Taiwan were eager to adopt the government's EFL graduation exam policy. Messick (1996) states that washback effects usually appear after the test has been used for a while. A longitudinal study would surely offer a clearer picture of the long term effects of this graduation exam policy.

Second, this study has relied on self-reported student data. As Nisbett and Wilson (1977) as well as Yu (2010) point out, such information is easily prone to expectancy bias. Subsequent research should include more classroom observational data and seek to corroborate student information with other data sources from teachers and school administrators. This should allow researchers to get a more accurate and dynamic picture of how washback patterns are perceived by different test stakeholders.

Third, one goal of introducing graduation requirements was to improve the ability of graduates to communicate effectively in English in the workplace. Little research has been conducted on the exploration of the possible washback effects of improved communicative competence. Although this would be very difficult for the researchers to measure, further research should pay more attention to this aspect of washback.

References

- Alderson, C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297.
- Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115–129.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
- Berry, V., & Lewkowicz, J. (2000). Exit-tests: Is there an alternative? *Hong Kong Journal of Applied Linguistics*, 5(1), 1949.
- Breen, M. P. (1984). Process syllabuses for the language classroom. In C. J. Brumfit (Ed.), *General English syllabus design* (pp. 47–60). Elmsford, NY: Pergamon Press & The British Council.
- 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, NJ: Lawrence Erlbaum.
- Chen, L. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the basic competence test in English* (Doctoral dissertation). Ohio State University.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. 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, V. H. (2006). English language education in Taiwan: A comprehensive survey. *Bimonthly Journal of Educational Resources and Research*, 69, 129–144.
- Chern, C. (2002). English language teaching in Taiwan. *Asia-Pacific Journal of Education*, 22(2), 97–105.
- Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62.
- 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* (Doctoral dissertation). National Taiwan Normal University, Taipei, Taiwan.
- Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. Wittrock (Ed.), *Handbook of research on teaching*, (3rd ed., pp. 255–296). New York: MacMillan.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London; New York: Routledge.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- 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, NJ: Lawrence Erlbaum.
- Fransson, A. (1984). Cramming or understanding? Effects of intrinsic and extrinsic motivation on approach to learning and test performance. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 86–121). London: Longman.
- Gall, J. P., Gall, M. D., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston: Allyn and Bacon.
- Gan, Z., Humphreys, G., & Hamp-Lyons, L. (2004). Understanding successful and unsuccessful EFL students in Chinese university. *The Modern Language Journal*, 88(2), 227–244.
- Gardner, R. C., & MacIntyre, P. D. (1993). An instrumental motivation in language study: Who says it isn't effective? *Studies in Second Language Acquisition*, 13, 57–72.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.
- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education*, 10(2), 169–207.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge University Press.
- Hayes, B., & Read, J. (2003). IELTS Test preparation in New Zealand: Preparing students for the IELTS academic module. In R. Tolloh (Ed.), *IELTS Research Report 4* (pp. 153–206). Canberra: IELTS Australia Pty Limited.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. W. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 97–111). Mahwah, NJ: Lawrence Erlbaum.
- Hsu, H.-F. (2010). *The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology* (Doctoral dissertation). University of York.
- IIBC (2005). *TOEIC Newsletter #89*. Tokyo: International Business Communication. Retrieved from http://toEIC.or.jp/toEIC_en/pdf/newsletter/newsletterdigest89.pdf
- Jou, C. (2010). *Perceptions of The Test of English For International Communication and its washback effects: A case study at a private university in Taiwan* (Master's thesis). Hsuan Chuang University, Taiwan.

- Lai, H. (2003). *Learning styles and personality types: Identification and comparison of hospitality students in Taiwan and United States* (Doctoral dissertation). Texas Tech University.
- Lam, H. P. (1994). Methodology washback and insider's view. *Bringing About Change in Language Education*, 83–99.
- Lane, S., Parke C. S., & Stone, C. A. (1998). A framework for evaluating the consequences of assessment programs. *Educational Measurement: Issues and Practice*, 17(2), 24–28.
- Menken, K. (2008). High-stakes tests as de facto language and education policies. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education* (2nd ed., Vol. 7, pp. 401–413). New York; London: Springer.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. doi: 10.1177/026553229601300302.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Pan, Y. (2009). Voices in the field: An interview with Jessica Wu. *SHIKEN: The Japan Association of Language Teaching, Testing & Assessment SIG Newsletter*, 13, 9–14.
- Pan, Y. (2010). *Consequences of Test Use: Educational and Societal Effects of English Certification Exit Requirements in Taiwan* (Doctoral dissertation). University of Melbourne, Australia.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Ceng, 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.
- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education*, 14(1), 51–74.
- Roeber, 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* (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.
- Schmidt, C. (2004). The analysis of semi-structured interviews. In U. Flick, E. V. Karloff & I. Steinke (Eds.), *A companion to qualitative research* (pp. 253–258). London: Sage.
- Shih, C. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135–162.
- Shohamy, E. (2001a). *The power of tests*. Harlow: Pearson Education Limited.
- Shohamy, E. (2001b). Democratic assessment as an alternative. *Language Testing*, 18(4), 373–391.
- 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 investigating the effects of test status on students' test preparation behaviours* (Doctoral dissertation). Hong Kong Polytechnic University, Hong Kong, China.
- 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.
- 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, UK: Cambridge University Press.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41–69.
- Wall, D., & Horak, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 1, the baseline study*. 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, NJ: Educational Testing Service.

- Watanabe, Y. (2004a). Teacher factors mediating washback. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129–146). Mahwah, NJ: Lawrence Erlbaum Associates.
- Watanabe, Y. (2004b). Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19–36). Mahwah, NJ: Lawrence Erlbaum.
- Watanabe, Y. J. (2001). Does the university entrance examination motivate learners? – A case study of learner interviews. In A. Murakami (Ed.), *Trans-equator exchanges: A collection of academic papers in honour of Professor David E. Ingram* (pp. 100–110). Akita: Akita University.
- 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.
- Yu, C. (2010). *Reliability of self-report data*. Retrieved from <http://www.creative-wisdom.com/teaching/WBI/memory.shtml>