

## ***Does Teaching to the Test Exist? A Case Study of Teacher Washback in Taiwan***

**Yi-Ching Pan**

*National Pingtung Institute of Commerce, Taiwan*

To enhance students' English proficiency and equip them with competitiveness in the workplace, an increasing number of universities/colleges in Taiwan have established English certification exit requirements, through which students must choose from and pass an array of English proficiency tests such as the GEPT, TOEIC, TOEFL, and IELTS in order to graduate. The goal of this study is to determine whether the exit requirements have brought "teaching to the test" to tertiary institutions or rather have acted to motivate teachers to integrate listening, reading, writing and speaking skills with lesson content to improve students' communicative competence. Data was collected from two groups of technical colleges throughout Taiwan, one group with and one group without exit requirements. The researcher distributed 160 teachers' questionnaires, interviewed 25 teachers, and observed about 50 English class sessions to appraise tertiary-level teaching washback. Findings indicated that the exit requirements have elicited a minor degree of changes in teaching. Teachers consider test factors and test-related activities the lowest priority in their selection of materials and pedagogy. However, a significant difference was found that teachers at schools with exit requirements have a higher consideration of test factors and employ more related activities than their counterparts.

**Keywords:** washback, English proficiency tests, teaching to the test

## **INTRODUCTION**

Washback is defined as the effects that tests have on teaching and learning (e.g. Alderson & Wall, 1993; Bailey, 1996; Hughes, 2003; Wall & Alderson, 1993). Teachers play an essential role in testing due to the fact that they have the tremendous power to inspire students to learn, to teach them language skills, and to model how to work with tests and test results (Bailey, 2005; Spratt, 2005). Since washback is a highly complex phenomenon that involves a number of variables, it should be investigated from multiple perspectives and be triangulated by different methods (Bailey, 1999; Cheng, 2001; Turner 2001; Wall, 2005; Watanabe, 2004a). However, most of the current washback studies utilized either qualitative or quantitative methods. Moreover, there is a lack of empirical evidence for washback; researchers (e.g. Alderson & Hamp-Lyons, 1996; Green, 2007; Wall & Alderson, 1993; Watanabe, 2004b) have called for a greater number of studies to be conducted to explore the washback of language tests more systematically. This study, therefore, aims to employ both qualitative and quantitative methods to generate a better understanding of the effects that tests have upon teaching practices in order to provide additional evidence that is currently lacking in the field of language testing. The two research questions employed in this empirical study are:

- (1) Do English certification exit requirements bring “teaching to the test” to higher education in Taiwan?
- (2) Do English exit requirements act as an impetus to motivate teachers to integrate the four language skills of listening, reading, writing, and speaking into lesson content to improve students’ communicative skills?

## **LITERATURE REVIEW**

In general, the majority of washback studies indicated that teachers tended to teach to the test, but this same research determined that any changes at the

classroom level as a result of testing are superficial rather than substantial. For example, Cheng (1997, 2004, 2005), Green (2007), Wall and Alderson (1993), and Wall (2005) found that while content changes because of the test, the way teachers instruct does not change to any significant degree. Cheng (2005) observed that the new 1996 Hong Kong Certificate of Education Exam in English encouraged teachers to pay more attention to the oral and integrated skills components. However, teacher talk remained the dominant activity in the classrooms observed. Wall (2005) discovered that teachers concentrated on the reading and writing skills that were tested on the new O Level English examination in Sri Lanka, particularly during the examination preparation period.

Similarly, Green (2007), Hayes and Read (2003, 2004), and Wall and Horak (2006) identified 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. Ferman (2004) discovered that due to the introduction of a new EFL oral test in Israel, teachers focused on the tested skills by coaching their students to do more oral-activity practice (e.g. interview, role-play, memorization of cue cards).

In contrast, other research determined that tests affected both how and what teachers taught, but not all teachers reacted the same way to the same test. Factors such as teachers' beliefs, educational backgrounds, and levels of experience certainly influence how washback works. For example, Shohamy (1993) and Shohamy et al. (1996) uncovered significant differences between experienced and novice teachers. The former tended to teach to the test and used only materials that would be included on the test, while the latter used different activities to teach oral language. Correspondingly, Lam (1994, p. 91) reported that more experienced teachers tend to be significantly more "examination-oriented" than their younger colleagues. However, this research did not explore why teachers reacted differently toward the same tests. Alderson and Hamp-Lyons (1996) and Watanabe (1996, 2004b) uncovered

significant variations in the way teachers instruct the same exam or exam skill. Some teachers resorted to explicitly traditional grammar-translation approaches to help students to learn the linguistic aspect of English, while others incorporated communicatively-oriented approaches to develop the real-life language abilities of their students.

Burrows (1998, 2001, 2004) also found that teachers responded to classroom-based assessment in different ways in the Australian Adult Migrant English Program. She categorized the teachers by the ways in which they reacted to this test as such: first, resisters; second, adopters; and third, partial adopters. The degree to which teachers' educational and philosophical beliefs were in accordance with the theoretical and educational foundations of the new assessment tool determined the type of teachers that they were. The most in accordance were the adopters, while the least in accordance were the resisters.

Regardless of whether changes occur to the content, methodology, or both, most teachers tended to "teach to the test" to some extent in order to help their students to familiarize themselves with the test content/format and to succeed on the assessment (e.g. Cheng, 2004; Ferman, 2004; Gu, 2007; Shohamy et al., 1996; Wall, 2005). Moreover, teachers' incomplete understanding of the nature of the exam, inadequate training, and professional backgrounds lead to unchanged methodologies because teachers simply do not know how to change, rather than because they do not want to (Cheng, 2001, 2005). For example, Qi's series of studies (2004, 2005, & 2007) reveal "cramming for testing" in teaching practices. The beliefs and attitudes that teachers held in regard to the immediate goals of assisting students to do well on a gate-keeping test, the NMET (National Maturation English Test in China), in conjunction with their own limited capacity to use the language effectively, contributed to their inability to bring about the positive changes intended by the test developers. One of the intended positive changes was a shift in English language teaching from a formal linguistic approach to a more communicative orientation.

As discussed above, the majority of studies indicated that teachers tended to teach to the test. However, several studies concluded that tests did not influence teaching to a significant degree. For example, Li's (2008) study demonstrated that there was only a minor impact on teaching brought about by the China English Tests (CET); teachers did not change either what or how they instructed to any significant degree. The factors for this weak level of washback, according to Li (2008), involved a number of variables, in terms of both test qualities (writing accounted for 15% of the CET total score, teachers possessed an insufficient knowledge of the CET and were concerned about its reliability) and other considerations (teachers' expertise and experience, beliefs, class size, and curriculum).

The preceding discussion makes it readily apparent that the effects that tests have upon teaching vary in degree of strength and in whether content, methodology, or both change. Because of the mixed findings drawn from these empirical washback studies, is it too early to conclude whether "teaching to the test" exists? Since most of the current washback studies utilized either qualitative or quantitative methods, this study, therefore, aims to employ both methods to determine whether English certification exit requirements have resulted in "teaching to the test" within the context of higher education in Taiwan, or if exit requirements have acted as an impetus to motivate teachers to integrate the four language skills of listening, reading, writing, and speaking into their lesson content to improve their students' communicative skills.

### **English Certification Exit 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. To that end, to enhance student workplace competitiveness, 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 nearly 90% in 2012.

The Ministry established a list of recommended tests to fulfill this requirement. The list includes the TOEIC<sup>®</sup>, TOEFL<sup>®</sup>, ITELTS, and two local tests, the General English Proficiency Test (GEPT) and the College Student English Proficiency Test (CSEPT). Table 1 lists the recommended ranking of the EFL exams recommended by the Taiwanese government in terms of two general proficiency levels. The GEPT is a five-level, four-skill general English proficiency examination commissioned by Taiwan's Ministry of Education (MOE) in 1999.

**TABLE 1**  
**The Taiwan Executive Yuan Department of Central Personnel**  
**Administration's 2006 Ranking of Some Current English Proficiency**  
**Tests in Terms of Two CEFR Levels**

CEFR Level	TOEFL <sup>®</sup>	TOEIC <sup>®</sup>	IELTS <sup>™</sup>	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+

## METHODOLOGY

To enhance both the validity and reliability of this research, this is a mixed-method study that employs questionnaires, interviews, and classroom observations to generate an understanding of how tests affect teaching in that each provides different but complementary data. Since it is more difficult to reach a large number of respondents through observations and interviews, it is essential to combine asking (interviews and questionnaires) and watching (observations) to explore the nature of washback effects (Cheng, 2005; Wall,

2005). By using these instruments, this study compared the differences between the schools with exit requirements (hereafter referred to as ECER schools) and those without exit requirements (hereafter referred to as non-ECER schools) to explore the effects that tests have upon teaching practices.

### Sample and Instruments

**TABLE 2**  
**Teacher Questionnaire Informant Profiles (N=160)**

		<b>EFL Teachers at ECER schools <i>n</i>=81</b>	<b>EFL Teachers at non-ECER schools <i>n</i>=79</b>
<b>Gender</b>	Male	25.9%	25.3%
	Female	74.1%	74.7%
<b>Academic position</b>	Professor	4.9%	1.3%
	Associate Professor	17.3%	8.9%
<b>Major</b>	Assistant Professor	3.7%	7.6%
	Lecturer	40.7%	82.3%
	TESOL	38.3%	46.8%
	Education	27.1%	18.9%
	Linguistics	1.2%	5.1%
	Applied Linguistics	9.9%	8.9%
	Others (mostly in Western Literature)	23.5%	20.3%
<b>Years of work experience</b>	1-5 years	19.8%	24.4%
	6-10 years	33.3%	32.1%
	11-15 years	2.5%	9.0%
	16-20 years	25.9%	28.1%
	More than 20 years	18.5%	6.4%

The teacher questionnaires attempted to gauge how teachers reputedly taught listening, reading, writing, and speaking skills and what materials they chose, with the end goal of discovering whether there are differences between teachers at the two groups of schools.

According to Table 2, 160 English teacher respondents from one-half (n=39) of the technical universities/colleges filled in the questionnaires. The ratio of teachers at ECER schools and non-ECER schools is nearly even. This even distribution is essential to this study because it was designed to explore the differences, if any, between the two groups of teachers, in terms of the effects of the exit requirement policy. Moreover, respondents had different educational backgrounds and years of work experience, and they worked in different locales.

Twenty-five teachers who expressed their willingness to be interviewed on the questionnaires were recruited for semi-structured interviews. Of these, 23 informants were instructors at ECER schools and two taught at non-ECER schools. In addition, two instructors worked at both groups of schools. You may refer to the interview guide in Appendix A.

Six English teachers were selected for their willingness to be observed in their English classes. The length and titles of the classes each teacher instructs, along with the number of students in each class, are illustrated below in Table 3. This was followed by a calculation of the average amount of time that the observed teachers devoted to each of the following categories over the course of the observations, as proposed by the COLT observation scheme (1995):

1. Skills (i.e. listening, speaking, reading, and writing) focused on;
2. Test-preparation activities;
3. Interaction between teachers and students (i.e. teacher-centered or student-centered interaction); and
4. Medium of instruction

The observation scheme used in this study appears in Appendix B.

**TABLE 3**  
**Summary of the Observed Teachers and Classes**

ID	Gender	Years of experience	ECER school	Non-ECER school	Title of English classes	Length of Observed classes	Number of students
C	M	<10 years		✓	General English	382.5min	63
			✓		Oral English	174.5 min.	43
B	M	<10 years		✓	General English	363.5 min.	44
			✓		English reading	316.5 min.	42
A	F	>20 years	✓		English reading	323.5 min.	50
D	F	>20 years		✓	General English	470.5 min.	56
E	F	<10 years	✓		Test-preparation	214 min.	40
F	M	<10 years	✓		Test-preparation	288.5 min.	40

### Data Analysis

SPSS 12.0 was used to analyze questionnaires. The Mann-Whitney U test was utilized for statistical differences with a probability of less than .05 for the ordinal 5-point Likert 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 should also facilitate any subsequent meta-analyses.

All the structured interviews were digitally recorded, transcribed, and then analyzed according to the analytical-strategy proposed by Schmidt (2004).

The classroom observations were based on the Communicative Orientation of Language Teaching Observation Scheme (COLT), developed by Spada and Froehlich in 1995.

## RESULTS

This section provides three major findings collected from teacher questionnaires, interviews, and classroom observations for the two research questions stated earlier:

- (1) Do English certification exit requirements bring “teaching to the test” to higher education in Taiwan?
- (2) Do the English exit requirements act as an impetus to motivate teachers to integrate skills of listening, reading, writing, and speaking in lesson content to improve students’ communicative skills?

### **Test-Related and Integrated-Skill Instruction Occurred with More Frequency in Regular English Classes at ECER Schools than at Non-ECER Schools, but not to a Significant Degree**

According to Table 4, two activities that exhibit statistically significant differences ( $p < .05$ ) between the two groups of schools concerned (1) the instruction of grammar, vocabulary, and phrases and (2) the instruction and practice of test-preparation material. These statistically significant differences, however, received somewhere between small and moderate effect sizes ( $d = 0.4$ ).

Another interesting finding is that traditional activities (e.g. instruction and explanation of textbooks, instruction of grammar, vocabulary, and phrases, explanation and practice of text exercises) were utilized slightly more in non-ECER schools, while various teaching activities such as group discussion, quizzes, practice of listening, writing, test-preparation, and language activities were employed at a marginally higher rate at ECER schools. Although teachers at ECER schools employed a wider array of teaching activities than their counterparts for at least a small portion of their class time (see items 9-12 in Table 3), it is important to point out that both groups devoted a lot of time to traditional activities such as textbook explanation or exercises for grammar and vocabulary.

**TABLE 4**  
**A Comparison of Class Activity Frequency at ECER and Non-ECER Schools**

Instruction Activities	ECER teachers			Non-ECER		
	N	M	SD	N	M	SD
<b>Test-related instruction activities</b>						
1. Giving quizzes	81	3.60	0.97	79	3.52	0.90
2. ☉-⊗* Instruction and practice of test-preparation material	80	3.04	0.85	79	2.72	0.96
<b>Traditional instruction activities</b>						
3. Instruction and explanation of textbooks	81	3.99	0.73	79	4.16	0.65
4. Explanation and practice of text exercises	81	3.91	0.64	79	3.99	0.69
5. ☉-⊗* Instruction of grammar, vocabulary and phrases	81	3.89	0.67	79	4.13	0.61
6. Listening to audio versions of text or radio broadcasts	81	3.51	0.88	77	3.43	0.99
<b>A variety of instruction activities</b>						
7. Practice of sentence or short essay writing	81	3.36	0.93	79	3.20	0.90
8. Group discussion/pair work	81	3.17	1.00	79	3.01	1.02
9. Role-play	81	2.89	1.05	79	2.90	0.98
10. Having students engage in language related activities	81	2.84	0.86	79	2.67	0.86
11. Watching English films	81	2.47	0.87	79	2.51	0.80
12. English songs	81	2.37	0.89	79	2.44	0.78

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

☉: small effect size ( $d \leq 0.3$ )    ⊗: moderate effect size ( $d = 0.5$ )

☉-⊗: small – moderate effect size ( $0.3 > d < 0.5$ )

In interviews, three major forms of test-driven instruction were reported by 22 interviewed teachers at ECER schools: 1) 17 teachers said they devoted some class time to practicing mock test questions, 2) 5 teachers indicated they did not instruct with test-preparation material in particular, but they

included test-related questions in mid-terms and final exams, or they gave more class quizzes, and 3) 10 teachers encouraged or required students to do test-related practice at self-access centers.

In observation (see Appendix C for details), teacher-centered instruction was the dominant form of instruction observed at both ECER and Non-ECER schools. Teachers usually stood in front of the students in a classroom and taught them directly from textbooks. Notable exceptions to this were in  $A^{ECER}$ 's and  $D^{non-ECER}$ 's classes. In  $A^{ECER}$ 's class, 36% of the interaction was student-centered; students were asked to do group discussion for collaborative writing tasks (e.g. creating a composition entitled "My own movie") and then do oral presentations for the assigned written tasks. 18% of the class time was student-centered in  $D^{non-ECER}$ 's class, where students were asked to conduct group discussion for the assigned written group work (e.g. a summary or a resume and cover letter) or to role-play with the teacher. This finding tends to be consistent with Watanabe's (2004b) and Burrow's (2004) studies, where teachers' individual factors rather than the test-driven policy impacted their teaching practices. However, considering the fact that students primarily used Mandarin for group discussion in  $A^{ECER}$ 's and  $D^{non-ECER}$ 's classes, these students did not seem to benefit from this communicatively oriented teaching approach. In addition, there was little teacher-student interaction in English when students did oral presentations in  $A^{ECER}$ 's class; the students usually read from their notes and the teacher commented in Mandarin.

#### **Test-related Materials were Adopted to a Higher Degree, but were not Dominant in Regular English Classes at ECER Schools**

Table 5 shows that test-related material had the second lowest mean ( $m=3.10$ ) from ECER teachers. Their counterparts had the lowest mean ( $m=2.68$ ), which represents a statistically significant difference ( $p<.05$ ). The difference resulted in a small-to-moderate effect size ( $d=0.4$ ). Although the majority of class materials used by both groups did not focus on test-related

content, the higher percentage (20% more) of such content at ECER schools suggests some washback in terms of teaching materials.

**TABLE 5**  
**A Comparison of Types of Texts Teachers Report Using**

Item	ECER teachers			Non-ECER teachers		
	N	M	SD	N	M	SD
1. Focusing on daily use of English	79	3.92	0.80	78	4.00	0.77
2. Focusing on the four skills	79	3.76	1.00	78	3.78	1.03
3. Self-made or supplementary materials	79	3.63	0.91	78	3.60	0.93
4. Authentic materials	80	3.38	0.95	79	3.43	0.98
5. Focusing on reading skills	79	3.27	1.06	78	2.91	1.06
6. ☉-⊗*	81	3.10	1.06	78	2.68	1.01
Test-related materials						
7. Beneficial for the work place	80	3.08	1.00	78	3.15	1.09

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

☉: small effect size ( $d \leq 0.3$ )

⊗: moderate effect size ( $d = 0.5$ )

☉-⊗: small –moderate effect size ( $0.3 > d < 0.5$ )

Responses from interviews showed that some ECER teachers use these test-preparation materials as supplementary materials, some as part of the textbooks, and some as the major texts in class. One thing that must be noted is that a few ECER teachers (n=4) mentioned the fact that the amount of test-preparation materials used in class was determined by the attitude of their administrators toward the exit requirement policy. For example, T45 stated that before the establishment of exit requirements, teachers had the freedom to choose the materials they found useful for their students. However, her

school implicitly suggested that English classes should focus on assisting students in passing the GEPT, so teachers chose GEPT test-preparation books and materials for the main text. Another English teacher (T43) claimed that the material he used in class was compiled on the basis of what is tested on the GEPT, “However, several years ago when fewer schools had set the exit requirements, the textbooks that I chose focused more on English literature.” Still another (T19) reported she had allocated appropriately 10% of her class time each week for test-preparation instruction and practice because her school requested that 50% of the mid-term and final exam questions be GEPT related.

Conversely, one teacher (T16) mentioned that test-preparation materials used to be adopted frequently in class when the school sought to improve its students’ pass rate. In addition, these materials were covered as part of the mid-term and final exams. However, since their pass rate has been sufficient, the school does not request that teachers adopt test-preparation materials. When asked if she now includes test-related content in class, she replied, “Not noticeably.”

As indicated in Table 4, exit requirements have brought about a higher use of test-preparation materials in class in general at these schools than at the schools of their counterparts, but the results of the teacher interviews show that the amount of these materials varies from teacher to teacher.

#### **Test-related Instruction was Found more in Optional Test-preparation English Classes at an ECER School**

The observation of six teachers over 4-5 weeks (see Appendix C) shows that although test-related instruction and practice in compulsory English classes at both ECER and non-ECER schools took up only a small amount of the class time, a slightly higher amount of time (2% of the average class time in contrast to 1.6%) was observed at the ECER school than at its counterpart.

However, “teaching to the test” occurred mostly in optional test-preparation classes that were observed in an ECER school. In those classes,

the negative washback effects mentioned in the literature were found: 1) instruction of test-taking skills rather than language learning activities (Wiseman, 1961; Noble and Smith, 1994), 2), a heavy reliance on test-preparation materials (Cheng, 2005; Wall, 2005), and 3) adoption of a memorization approach, where students were encouraged to do mechanical drill practice and memorize a number of vocabulary items, phrases, and even model answers (Alderson, 1990). However, these test-preparation classes were optional and the number of such classes was a minority.

Explicit instruction of test-preparation was apparently observed much more in the optional test-preparation classes at the ECER school, which were dominated by the test. The textbooks used for this class were mainly for GEPT preparation. The two observed teachers followed what was written in the text and provided their students with an abundance of listening and speaking practice related to what is tested on the elementary GEPT, such as practicing answering the oral questions under the time limit. Test-taking strategies were often mentioned in class to remind the students what they could do to select the correct answers or to answer oral questions in ways that may be more favored by test raters. Decontextualized learning was often observed in the test-preparation classes in which the teacher played an active role, and the students were engaged in practice-oriented activities, such as listening to the recording and choosing the correct answer on a picture, repeating after the teacher, and practicing the possible alternative answers to the oral questions.

This finding suggests that English exit requirements have more impact on teaching practices in optional test-preparation classes than in regular English classes at ECER schools.

## **DISCUSSION**

As found in the literature (e.g. Cheng, 2005; Wall, 2005; Wall & Alderson, 1993; Watanabe, 2004a), “the process of washback being generated is

mediated by numerous factors” (Watanabe, 2004a. p. 19). The weak level of test effects observed in ECER schools can perhaps be attributed to both testing and mediating factors, as described below.

Regarding the test factors, most English certification tests are in the format of multiple-choice questions, which resulted in some teachers contending that test scores do not really demonstrate one’s level of English proficiency. In addition, the tasks that some English certification tests give are not applicable to daily life, so it is not useful for students to memorize vocabulary without knowing how to use it within a context or to do monologue practice with a machine, like in the oral section of the GEPT. Thirdly, the test content in some English certification tests not only assesses language proficiency but also something other than that, such as math skills. In other words, these teachers were concerned with construct-irrelevance, introduced by Messick (1989), which may influence the validity of the test. In light of these concerns about test quality, more positive test effects may be generated if test developers attempted to develop a test that could “encourage teachers to change their way of teaching” (Wall, 1997. p.724). Likewise, Shepard (2000) and Robinson (2000) propose that test assessment practices should align with current curriculum theories and pedagogical theories. Stomp (2008) also supports their idea and demonstrates in his study that negative test effects on teaching and learning usually occur when the gap between testing and pedagogy is not addressed.

Regarding mediating factors, the amount of test effects may depend upon teachers’ perceptions of English certification exit requirements, students’ proficiency, administrators’ attitudes toward this policy, the mismatch between curricula and test contents, and situational factors such as class size and additional number of hours of English classes. For example, teachers’ educational backgrounds and teaching beliefs led to their opposition to crammed education, where test-preparation practice was the class focus. Instead, they believe that it is more essential to help students to build a solid foundation of the basic four skills, and once they are competent in these skills, they should be able to pass the test, particularly the GEPT, which many

students take to meet the requirement. Even though the GEPT assesses the four skills, because of the large size of classes and students' low proficiency, many interviewed teachers claimed that they had difficulties incorporating a significant amount of speaking and writing practice in class. Moreover, in some ECER schools, fixed syllabi left teachers with very little time to assist students with test-preparation. The offering of extra test-preparation classes also made teachers feel that there was no urgent need to include many test-preparation materials in class.

## CONCLUSIONS AND IMPLICATIONS

By utilizing a variety of instruments including questionnaires, interviews, and classroom observations, this study found that the exit requirement policy did not appear to increase the number of teachers teaching to the test, a phenomenon often reported in an examination-oriented society like Taiwan (Chern, 2002; Huang, 2001; Lai, 2003; Tsai & Tsou, 2009). Moreover, this exit requirement policy did not encourage instruction of productive skills or communicative-oriented approaches to a large extent because merely a minority of the ECER teachers, less than one third, claimed that they frequently did these activities. However, classroom observations show that optional after-school test-preparation classes at an ECER school adopted more explicit test-preparation activities, compared to regular English classes at both ECER and non-ECER schools. Test practice instruction and activities in the test-preparation classes that were dominated by tests included 1) the adoption of test-preparation materials as the main source for instruction, 2) the encouragement of memorization of vocabulary, phrases, sentence structures, and key answers to questions often appearing on the test, 3) the practice of mock tests in a timed condition, and 4) the explanation of test-taking strategies and the preferences and expectations of raters.

To conclude, at the time of this study, teachers at ECER schools tended to focus more on test-related skills, adopt more test-preparation instruction, and

use more varied communicatively-oriented activities. If exit requirements continue to be promoted as an educational policy by the MOE to enhance students' English proficiency, in addition to regular teaching, teachers must work together to determine how to help students to do well on tests without necessarily teaching to the test. This might involve increasing students' motivation and interest in learning English through the use of carefully designed teaching programs and choice of appropriate materials. The alignment of curricula with test content (e.g. the GEPT) may also be one possibility teachers can consider so that they will be able to focus both on receptive and productive skills in class. Professional development should be provided to assist them to build an appropriate curriculum that suits students' interests and skill levels, and also integrate what is covered on English certification tests.

In addition, the offering of test-preparation instruction could eliminate the gap between teachers' actions and students' desires and make students feel confident when preparing for tests. This is especially true for low-achieving students, who often claimed at the interviews that they did not receive a sufficient amount of test-preparation instruction and practice. Thus, they had little idea about what and how to study for the test, which might be one of the factors leading to the weak amount of washback on learning. According to Roderick and Engel (2001), sufficient test-preparation is beneficial for students. They contend that explicit instruction should be given to low-achieving students to demonstrate to them how and what to study.

In order to generate a better understanding of washback variability on teachers and add valuable evidence to this research field, future studies should include an analysis of instructors' years of experience and their English proficiency levels as factors to determine whether they are related to their teaching practices in the classes with and without test-driven requirements.

## **THE AUTHOR**

*Yi-Ching Pan* has a Ph.D. in Applied Linguistics from the University of Melbourne and is currently an Associate Professor at the National Pingtung Institute of Commerce in southern Taiwan. Since 2009 she has been on the editorial board of the International Journal of Educational and Psychological Assessment and in January 2011, she edited a special edition of that publication on foreign language assessment. Her research interests include second language acquisition and language testing.  
Email: yichingpan@yahoo.com.tw

## REFERENCES

- Alderson, J. C. (1990). The relationship between grammar and reading in English for academic purposes test battery. In D. Douglas & C. Chappelle (Eds.), *A new decade of language testing research: Selected papers from the Annual Language Testing Research Colloquium* (pp. 203-219). Alexandria, VA: Teachers of English to Speakers of Other Languages.
- 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.
- 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.
- Bailey, K. M. (2005, March). *Teachers and tests: Promoting positive washback*. Paper presented at the TESOL Arabia Conference, Dubai, United Arab Emirates.
- Burrows, C. (1998). *Searching for washback: An investigation into the impact on teachers of the implementation into the adult migrant English program of the certificate in spoken and written English*. Unpublished doctoral dissertation, Macquarie University, Sydney, Australia.
- 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.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L. (2001). Washback studies: Methodological considerations. *Curriculum Forum*, 10(2), 17-32.
- 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, NJ: Lawrence Erlbaum.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Chern, C. (2002). English language teaching in Taiwan. *Asia-Pacific Journal of Education*, 22(2), 97-105.
- 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.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press
- Gu, X. (2007). *Positive or negative - An empirical study of CET washback*. 重慶: 重慶大學出版社 [ChongQing: Chong Qing 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.
- Huang, C. (2001). Using authentic materials in Taiwan's English classroom: The best way to enhance students' English proficiencies and intercultural competence. *Journal of National Taipei University of Technology*, 34(1), 225-234.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

- Lai, H. (2003). *Learning styles and personality types: Identification and comparison of hospitality students in Taiwan and United States*. Unpublished doctoral dissertation, Texas Tech University.
- Lam, H. P. (1994). Methodology washback and insider's view. *Bringing About Change in Language Education*, 83-99.
- Li, Hong. (2008, June). *Are teacher teaching to the test? A case study of the College English Test (CET) in China*. Poster presented at 30<sup>th</sup> Annual Language Testing Research Colloquium, Hang Zhou, China.
- 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.
- Mesick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Noble, A. J., & Smith, M. L. (1994). *Measurement-driven reform: Research on policy, practice, repercussion*. CSE Technical Report 381, Tempe, AZ: Arizona State University, CSE.
- 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.
- 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.
- Robinson, S. (2000). "Miles to go...": Assessment and secondary English. In B. R. C. Barrell & R. F. Hammett (Eds.), *Advocating change: Contemporary issues in subject English* (pp. 254-280). Toronto: Irwin Publishing.
- Roderick M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low achieving pupils to high stakes testing, *Educational Evaluation and Policy Analysis*, 23, 197-228.
- Schmidt, C. (2004). The analysis of semi-structured interviews. In U. Flick, E.V. Kardoff & I. Steinke (Eds.), *A companion to qualitative research* (pp. 253-258). London: Sage.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. *Washington, DC: National Foreign Language Centers Occasional Papers*.

- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13(3), 298-317.
- Spada, N., & Froehlich, M. (1995). *COLT: Communicative orientation of language teaching observation scheme, coding conventions and applications*. Sydney, NSW: Macquarie University, National Center for English Language Teaching and Research.
- 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.
- Stomp, D. H. (2008). Harming not helping: The impact of a Canadian standardized writing assessment on curriculum and pedagogy. *Assessing Writing*, 13, 180-200.
- Tsai, Y., & Tsou, C. (2009). A standardised 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.
- Turner, C. (2001). The need for impact studies of L2 performance testing and rating: identifying areas of potential consequences at all levels of the testing cycle. In A. Brown et al. (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies, studies in language testing 11* (pp. 138-149). Cambridge: UCLES/Cambridge University Press.
- Wall, D. (1997). Impact and washback in language testing. In D. Corson (Ed.), *Encyclopedia of language and education. Language testing and assessment* (Vol. 7, pp. 291-302). Dordrecht; Boston: Kluwer.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge: 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: Phrase 1, the baseline study*. Princeton, NJ: Educational Testing Service.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333.
- Watanabe, Y. (2004a). 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. (2004b). 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, Publishers.
- Wiseman, S. (Ed.). (1961). *Examinations and English education*. Manchester, England: Manchester University Press.

## **APPENDIX A**

### **Interview Guide for Teachers**

#### **(A) Perceptions of English certification tests**

1. What do you think about the English certification tests?
2. Do you think the GEPT can assess students' communicative competence?
3. Do you think the GEPT can demonstrate students' English abilities?
4. Which English certification test did your students choose to take?
5. Which English certification test would you recommend your students to take?

#### **(B) The differences among teachers at ECER schools in regard to lesson content, teaching methodology, and teaching materials**

1. Since your school has the English certification exit requirements, have made any changes in your teaching and teaching materials to address them or do you do what you used to do?
2. What percentage of class time is devoted to test-related instruction or practice?
3. In choosing materials for class, do you choose test-related materials? Or is test-related material not a major consideration for choosing a textbook?
4. What test-related instruction has been done to comply with the requirements?
5. Which skills are more focused on than others? And why?
6. The major goal of the exit requirements is to help to enhance students' proficiency and increase the number of students who are awarded the certificates. In light of this, why wouldn't you devote a major portion of your class time to practice of test-preparation materials?
7. Since there are no such exit requirements for non-English majors, do teachers instruct on anything that they find important or is there an

agreement on the objectives of the curriculum among teachers at your school?

8. Your school has not set any requirements. As any certificates are considered advantageous these days, have you included any test-related content in your teaching?

**(C) The measures that the school should provide with teachers to address the exit requirements**

1. Are there additional measures or resources to address the requirement?
2. What measures do you want your school to provide to teachers to reach the goals set for the exit requirements?

**(D) The positive and negative test effects brought about by the English certification exit requirements**

1. What positive or negative remarks have you heard from the students about the requirements?
2. What positive or negative effects have been brought about on your teaching?
3. Have the English certification exit requirements exerted some sort of pressure, such as an increase of teaching load?
4. Have you seen any changes or differences in students' motivation or desire to learn English?
5. You have experience teaching both groups of students, one with and one without the requirement. Have you found a difference in terms of their motivation to learn English?
6. Some teachers find that the requirement is not beneficial to students with low-level proficiency. What do you think about this?
7. Do you think the requirement might prevent the students with low proficiency from learning English instead?

8. What effects do you think the requirements have on teaching, such as the integration of four skills or leading to teaching to the test instead?
9. Are teachers given control in terms of what should be taught in class?
10. Some teachers thought the requirement could lead to teaching to the test, that the class would become boring, or cultural issues could not be covered in class. Have you also experienced these negative effects in your teaching?

## **APPENDIX B**

### **Observation Scheme**

#### **Observation analysis scheme**

**ID of the teacher observed:** \_\_\_\_\_

Recording number	
English as medium of instruction	
Interaction (teacher-student)	
Activities	
Listening	
Reading	
Speaking	
Writing	
Test-preparation	
Laughter	
Class materials	
Class time (min)	

## APPENDIX C

Summary of Classroom Observations at Two Schools Indicating the Time Allotted to the Skills Focused on, Activities for Test-Predation, and Teacher-Student Interactions

English compulsory classes at an ECER school	A <sup>ECER</sup>	B <sup>ECER</sup>	C <sup>ECER</sup>
<b>Class</b>	Reading (323.5min)	Reading (316.5min)	Oral (174.5min)
<b>Interaction</b>	1. teacher-centered 64% 2. student-centered 36%	teacher-centered	teacher-centered
<b>English as medium of instruction</b>	2.1%	7%	30.5%
<b>Skills focused on</b>	Reading: 52% Writing: 31.6%	Reading: 67% Listening: 17.4%	Listening: 31.8% Speaking: 43.5%
<b>Test-preparation</b>	1. mock-tests held once a month (not observed) 2. 10 hours of after-class practice at self-access center (not observed) 3. no explicit instruction of test-taking strategies or test-related practice	5.1% 1. doing textbook listening and writing exercise similar to the format of the GEPT 2. encouraging students to prepare for the GEPT 3. self-reading of two GEPT preparation textbooks (not observed)	1% for an introduction to the GEPT: a. test format b. test dates c. the importance of a rich vocabulary bank and good grammar knowledge for receiving high marks
<b>Activities</b>	1. giving lectures 2. choral repetition 3. grammar check 4. comprehension check <b>5. group discussion</b> <b>6. oral presentations</b> <b>7. team work for written assignments</b>	1. giving lectures 2. choral repetition 3. reading the passage 4. mechanical oral pair practice	1. giving lectures 2. choral repetition 3. reading the passage 4. grammar check 5. comprehension check 6. asking questions (but few engaged in discussion)
<b>Laughter</b>	1 laughter episode in each observed session on average	1 laughter episode in each observed session on average	1 laughter episode in each observed session on average

Notes: 1. A<sup>ECER</sup>, B<sup>ECER</sup>, C<sup>ECER</sup> are 3 teachers at an ECER school while D<sup>non-ECER</sup>, B<sup>non-ECER</sup>, C<sup>non-ECER</sup> are teachers at a non-ECER school. B and C teach both at ECER and non-ECER schools.  
2. Student-centered interaction activities are listed in bold.

**APPENDIX C**  
**Continued**

English compulsory classes at a non-ECER school	D <sup>non-ECER</sup>	B <sup>non-ECER</sup>	C <sup>non-ECER</sup>
<b>Class</b>	Reading (470.5min) teacher-centered 82% student-centred 18%	Reading (363.5min)	Reading (382.5min)
<b>Interaction</b>		teacher-centered	teacher-centered
<b>English as medium of instruction</b>	5.7%	9.4%	5.7%
<b>Skills focused on</b>	Reading: 45.4% Writing: 26%	Reading: 73.4% Listening: 23.3%	Reading : 44.7% Writing : 20%
<b>Test-preparation</b>	encouraging students to take the GEPT	1.occasional instruction of test-taking strategies for listening 2.50% of final questions related to the listening section of the GEPT	<b>4.8% for an introduction to the GEPT:</b> a. test format b. test dates c. educational resources provided by the school for test preparation d. information as to where to purchase test-preparation materials
<b>Activities</b>	1. giving lectures 2. choral repetition 3. reading the passage 4. grammar check 5. comprehension check <b>6. group discussion</b> <b>7. role-play</b> <b>8. team work for written assignments</b>	1. giving lectures 2. choral repetition 3. reading the passage 4. grammar check 5. comprehension check 6. asking questions (but few engaged in discussion)	1. giving lectures 2. choral repetition 3. reading the passage 4. grammar check 5. comprehension check
<b>Laughter</b>	1 laughter episode in each observed session on average	1 laughter episode in each observed session on average	1 laughter episode in each observed session on average

Notes: 1. A<sup>ECER</sup>, B<sup>ECER</sup>, C<sup>ECER</sup> are 3 teachers at an ECER school while D<sup>non-ECER</sup>, B<sup>non-ECER</sup>, C<sup>non-ECER</sup> are teachers at a non-ECER school. B and C teach both at ECER and non-ECER schools.  
2. Student-centered interaction activities are listed in bold.

## APPENDIX C

### Continued

Test-preparation classes at an ECER school	E <sup>ECER</sup>	F <sup>ECER</sup>
<b>Class</b>	for the listening/speaking sections of the elementary GEPT (214 min)	for the listening/speaking sections of the elementary GEPT (288.5min)
<b>Interaction</b>	teacher-centered	teacher-centered
<b>English as medium of instruction</b>	1%	3.2%
<b>Skills focused on</b>	Listening: 61.6% Speaking: 12.6%	Listening: 8.6% Speaking: 12.2% (87.8% of the class time was spent on the instruction of speaking skills, but 12.2% was devoted to actual speaking practice such as choral repetition and answering oral test questions.)
<b>Test-preparation</b>	Explicit test-preparation instruction and practice: 1. test-taking strategies for listening 2. test-question practice 3. test-question explanation 4. GEPT practice questions often appeared 5. use of test preparation materials & self-access center	Explicit test-preparation instruction and practice: 1. test-taking strategies for listening & speaking 2. test-questions practice 3. test-questions explanation 4. preference of raters 5. GEPT practice questions often appeared
<b>Activities</b>	1. giving lectures 2. choral repetition 3. reading the passage 4. practicing tests 5. comprehension check	1. giving lectures 2. choral repetition 3. reading the passage 4. practicing tests 5. comprehension check
<b>Laughter</b>	1 laughter episode in each observed session on average	5.2 laughter episode in each observed session on average

Notes: The percentage of time taken for the medium of instruction, interaction, test-preparation, and skills focused on, was calculated by counting the actual time for each session and then dividing this by the total duration of the session. Finally, the percentage of time taken for each session was added up and then divided by the number of sessions. For example, 8 sessions of A's classes were observed. 48% of the first session was spent on the instruction of reading skills, 77% for the 2nd, 22% the 3rd, 8% the 4<sup>th</sup>, 70% the 5<sup>th</sup>, 65% the 6<sup>th</sup>, 64% the 7<sup>th</sup>, and 63% the 8<sup>th</sup>. So the average percentage of time spent on reading skills in A's class was  $(48+77+22+8+70+65+64+63)/8=$