

How tests change teaching: A model for reference

CHIH-MIN SHIH

National Institute of Education, Nanyang Technological University

ABSTRACT: The purpose of this study was to investigate the washback effects of the General English Proficiency Test (GEPT) on English teaching in two applied foreign language departments in Taiwan. One had prescribed its GEPT requirement to its day-division students whereas the other had not. Overall, the GEPT did not induce a high level of washback on teaching in either department. Only courses which were linked to the departmental GEPT policy and whose objectives were to prepare students for the test were significantly affected. Results suggest that micro-level contextual factors (for example, the objectives of the course) and teacher factors had a greater impact on teachers' instruction. Finally, on the basis of current understandings of washback, I propose a new, tentative model to portray the washback of tests on teaching.

KEYWORDS: Washback, English teaching, EFL instruction.

INTRODUCTION

Washback¹ or backwash has been defined as “a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large” (Hughes, 2003, p. 53). For decades, testing has been purposely adopted by politicians, policy-makers and educators to bring about an impact on teaching, learning and other aspects of educational reform. Although being universally used for various purposes, testing is considered by scholars and researchers to induce mostly detrimental washback on teaching. For example, in general education, researchers have claimed that high-stakes testing might trigger a myriad of unethical test preparation practices or motivate teachers to manipulate students' test scores (for example, Falk, 2002; Haladyna, Nolen, & Haas, 1991; Jones, Jones, & Hargrove, 2003), or it might cause teachers to teach to the test (for example, Brennan, 2001; Janesick, 2001; Kohn, 2000; Rich, 2003). Similarly, tests are often deemed deleterious in language education, as Alderson and Banerjee (2001) have pointed out in a review article.

The above-mentioned beliefs in the effects of tests, however, have seldom been empirically corroborated in the area of language testing. It was not until Alderson and Wall (1993), who proposed their Washback Hypothesis as the foundation for further research, that researchers undertook washback studies. To date, researchers have paid most of their attention to the washback of tests on four domains of teaching practice: (1) content of teaching, (2) teaching methods, (3) assessment methods, and more broadly (4) overall teaching style, classroom atmosphere and teachers' feelings

¹ Washback is also known as backwash (Cheng & Curtis, 2004; Wall, 1997). Because washback is more universally used than backwash in the field of applied linguistics (Alderson & Wall, 1993; Cheng & Curtis, 2004; Green, 2007; Hawkey, 2006), I will use washback henceforth.

toward the test. Among the four domains, teaching content was always found to be altered by tests (for example, Alderson & Hamp-Lyons, 1996; Cheng, 2005; Ferman, 2004; Hawkey, 2006; Stecher, Chun, & Barron, 2004). The only exception was Shohamy, Donitsa-Schmidt and Ferman's (1996) study, in which they found that the Arabic Second Language (ASL) test in Israel did not bring about any washback on the content of teaching owing to its low-stakes nature and the low status of Arabic in Israel.

In contrast to the universal presence of tests' washback on the content of teaching, teaching methods were not changed by tests to a great extent. They were shown to be altered in some studies (for example, Alderson & Hamp-Lyons, 1996; Ferman, 2004; Hawkey, 2006; Stecher *et al.*, 2004), but not in others (for example, Cheng, 1997, 1998, 1999; Wall & Alderson, 1993). In addition, research has shown that tests influence how teachers administered tests (Wall & Alderson, 1993). As for the impact of the tests on overall teaching style, classroom atmosphere and teachers' feelings toward the test, research results indicate that there was more teacher talk in test-preparation courses (Hawkey, 2006). In addition, tests made courses more structured but less entertaining and dynamic (Hawkey, 2006). They also imposed anxiety and fear on teachers owing to their high-stakes status (see Ferman, 2004). It was also noted that washback happened for some teachers but not for others (for example, Burrows, 2004; Watanabe, 1996).

It can be concluded that different language tests induce a wide spectrum of washback on the content of teaching, teaching methods, and teacher-made assessment. Many factors, which I divide into the following three categories, were reported by empirical studies to affect the degree of washback on teaching: contextual factors, test factors and teacher factors. Contextual factors include management within schools (Wall & Alderson, 1993), the size of the class (Alderson & Hamp-Lyon, 1996), the objectives of the course (Hayes & Read, 2004), the timing of the course that is offered (Shohamy *et al.*, 1996; Watanabe, 1996), the professional support teachers receive from the school or test designers (Hawkey, 2006), resistance from other subject teachers owing to class schedules (Hawkey, 2006), and the variation of students' abilities in the class (Hawkey, 2006).

Test factors comprise the stakes of the test (Alderson & Hamp-Lyons, 1996; Shohamy *et al.*, 1996), the extent to which the test is counter to current teaching practice (Alderson & Hamp-Lyons, 1996), the extra administrative work entailed by the test (Hawkey, 2006), the language skills tested (Shohamy *et al.*, 1996), and the status of the language tested (Shohamy *et al.*, 1996). Teacher factors consist of teachers' abilities in the language they teach (Qi, 2007), teaching experience (Watanabe, 1996), teacher training processes or levels of professional training (Green, 2006; Shohamy, 1993; Wall & Alderson, 1993; Watanabe, 1996), teachers' beliefs about effective teaching, learning and test preparation (Alderson & Hamp-Lyons, 1996; Green, 2006; Wall & Alderson, 1993; Watanabe, 1996), learning experience (Watanabe, 1996), teachers' concerns for students' proficiency levels (Watanabe, 2004), the degree of teachers' familiarity with a range of teaching methods (Watanabe, 2004), teachers' perceptions of test importance (Shohamy *et al.*, 1996), perceptions of the test qualities (Shohamy *et al.*, 1996), other obligations (for example, teaching obligations in other schools) (Wall & Alderson, 1993), the degree of teachers' commitment to the teaching profession (Wall & Alderson, 1993), and teachers' willingness and capability

to innovate (Wall & Alderson, 1993).

The factors which I have just enumerated demonstrate that washback is a complicated phenomenon. To better portray the washback of tests on teaching, Burrows (2004) analyzed qualitative data from classroom observations to investigate teachers' responses to a newly introduced test. She found that their reactions to the test followed certain patterns instead of a single, uniform response or individual, different responses. Therefore, she proposed her curriculum innovation model, in which she pointed out that a new test, interpreted and shaped by teachers' beliefs, assumptions and knowledge, would lead to patterns of response in teachers' instruction. The curriculum innovation model, according to Burrows, established its basis on the concept that washback was a form of educational change, so behavioral models propounded for other educational changes could also be adopted to explain washback.

THE PRESENT STUDY

The present study investigated the washback of the General English Proficiency Test (GEPT) on teaching and learning² in Taiwan. This issue deserves investigation because the GEPT is a national English proficiency test, whereas the tests investigated in previous washback studies were national matriculation English tests (for example, Cheng, 2005; Qi, 2007; Shohamy *et al.*, 1996; Wall, 2005; Watanabe, 2004), academic English tests (Alderson & Hamp-Lyons, 1996; Green, 2007; Hawkey, 2006; Hayes & Read, 2004), classroom-based assessment (for example, Burrows, 2004), and second-language tests in high schools (Shohamy *et al.*, 1996). Because the GEPT has distinct purposes and its test-takers are from all walks of life across different age brackets, investigating its washback might have the potential to shed light on different dimensions of washback mechanisms. Before elaborating on the research methods adopted, I will briefly describe the GEPT.

The GEPT

Commissioned by the Taiwan Ministry of Education (MOE), the Language Training and Testing Center (LTTC) in Taiwan introduced the GEPT to provide a fair and reliable test for Taiwanese English-learners at all levels. It was hoped that this indigenous test would promote life-long learning and encourage English study (LTTC, n.d.-a).

The GEPT was the first large-scale examination of English in Taiwan to incorporate listening, speaking, reading and writing tests. It includes the elementary, intermediate, high-intermediate, advanced and superior levels. Except for the superior level, which integrates all tested skills into a four-hour test, the rest of the test levels have two stages (LTTC, n.d.-a.). Listening and reading abilities are examined in the first stage, and speaking and writing skills in the second stage (LTTC, n.d.-b, n.d.-c, n.d.-d, n.d.-e). Passing the first phase of a specific level is the prerequisite for registering for its second stage (LTTC, n.d.-a). Test takers who pass both phases of a specific level can

² The research report on the impact of the GEPT on English learning has been published in Canadian Modern Language Review (see Shih, 2007). This paper focuses solely on its impact on teaching.

receive a certificate of achievement from the LTTC (LTTC, n.d.-a).

Since its debut in 2000, the GEPT has sparked an unprecedented “whirlwind” in English testing in Taiwan. It has been used as the indicator of the English abilities of teachers and students by the former government as part of its governmental four-year plan for 2005-2008 (MOE, 2005, March 9). Now, the number of registered test-takers has reached 2.7 million (LTTC, n.d.-a). Numerous universities in Taiwan have chosen the GEPT for admission to university studies, as a placement test, and as the benchmark for graduation. It is also universally adopted for various purposes by the private sector and governmental institutions.

METHODS

When it comes to investigating washback, researchers tend to compare two courses, tests, or other potential variables, to gain a deep understanding of this educational phenomenon. For example, Shohamy (1993) and Shohamy *et al.* (1996) compared the impact of the ASL and English Foreign Language (EFL) tests. Alderson and Hamp-Lyons (1996) analyzed the TOEFL³ and non-TOEFL courses taught by the same instructor. Hayes and Read (2004) focused on two IELTS⁴ preparation courses; one was more test-focused, and the other leaned toward English for academic purposes.

I was inspired by these research designs, so I selected the applied foreign language departments of a university of technology⁵ (University A) and an institute of technology (University B). At the time of my data collection, the former did not impose a GEPT requirement. The latter required day-division students, both in its 5-year, junior college program and 4-year college program, to pass the first stage of the GEPT’s intermediate level or the school-administered make-up examination one year before the end of their academic studies.

With the exception of this difference, the two universities were akin to each other in the following ways. Firstly, both universities were private and located in urban areas of central Taiwan. University A is situated in one of the major cities in central Taiwan and University B in the largest metropolis in central Taiwan. Secondly, students at both universities were primarily from central Taiwan, although their residences might skew to the city and its adjacent areas where the university is situated. Thirdly, students at both universities had similar English proficiency and mostly took the elementary or intermediate levels of the GEPT. Fourthly, both universities’ applied foreign language departments had a short departmental history; university A’s was inaugurated in 1998 and University B’s in 2000. Fifthly, the participating universities were in the vocational education system, so they were able to complement previous washback studies on the GEPT, which focused on national research universities (Chuang & Hsu, 2006; Vongpumivitch, 2006) and public, senior-high schools (Lai, 2003; Wu & Chin, 2006).

³ TOEFL stands for Test of English as a Foreign Language.

⁴ IELTS is the International English Language Testing System.

⁵ The institute of technology became a university of technology soon after I completed this study.

Pseudo-nym	Gender	Age	Educational background	Nationality	Observed course		
					Course title	Division	No. of students
University A							
Carmella	Female	Early 50s	Ph.D. candidate in linguistics	Canadian	English Writing	Weekend	40
Sheila	Female	Early 40s	Ph.D. in agricultural education	Taiwanese	English Listening and Speaking	Day	30
University B							
Angela	Female	Late 30s	M.A. in TESOL	Taiwanese	GEPT-Preparation	Day	70
Don	Male	Early 40s	M.A. in English literature	American	English Listening and Speaking	Day	55
Paul	Male	Late 20s	Ph.D. student in educational policy and administration	Taiwanese	Vocabulary and Reading	Night	40

Table 1. Information on the participating instructors and their observed courses

In each department, the department chair, two or three teachers⁶, and 14 or 15 students participated in this study. In recruiting teacher participants, gender, age, nationality, educational background, the link of the course to the university's GEPT policy (for example, Don's Listening and Speaking course⁷), the content of the course which the prospective teacher was teaching (that is, the English skills taught were tested in the GEPT), and the division (that is, day, night or weekend divisions) in which the course was offered were adopted as the criteria to select teachers with diverse backgrounds. The recruitment of student participants was twofold. Some were chosen from the classes which I observed, so that I could triangulate the results of my observations with findings from teachers' and students' interviews. Others were selected from different classes in day, night or weekend divisions to cast light on the GEPT's impact on the teaching of those teachers whom I did not have an opportunity to observe. I used the following methods for this research:

Observations

I started my data collection by conducting classroom observations. The duration of instruction for each course was two hours per week over the 18 weeks of a semester. Because more themes emerged from Angela's GEPT-Preparation course, I observed for all course meetings. For the rest of the courses, I selected only eight weeks to conduct classroom observations. When selecting the specific weeks for classroom observation, I considered the time of the GEPT's administration (prior to or after the test), and different times of the semester (beginning, middle and end of the semester).

⁶ See Table 1 for information about participating teachers and their observed courses.

⁷ Day-division students in the applied English department of University B were required to pass the first stage of the GEPT's intermediate level before they completed Don's Listening and Speaking course, which was offered one year before they graduated. Otherwise, they were given the chance to take the make-up examination administered by their department prior to the end of the semester. If they failed both tests, they would automatically fail their Listening and Speaking course according to the departmental GEPT policy.

When conducting classroom observations, I took notes on a form that I had designed (see Appendix A). In addition, because GEPT-relevant data emerged more frequently in the GEPT-Preparation course at University B, I audio-recorded this course with a digital audio-recorder as an auxiliary measure for data collection.

Interviewing

I interviewed all participants once and recorded the interviews with a digital audio-recorder. Interviews with the teachers were conducted when scheduled classroom observations were about to be completed. Prior to the interview, each participant chose a pseudonym used in the research and filled out a profile questionnaire. Thereafter, interviews were guided by the interview schedules I devised prior to the research (see Appendix B). I interviewed the department chair to investigate the departmental history and policies on English requirements. Participating teachers were interviewed about their teaching and their views of the GEPT. Students were interviewed (1) to triangulate with the accounts of teachers who were interviewed and (2) to examine whether other teachers who were not interviewed and observed prepared students for the GEPT. Interviews with department chairs, teachers and students lasted about 30 to 40 minutes, 25 to 55 minutes, and 20 to 60 minutes, respectively. Except for two expatriate teachers who were interviewed in English, other participants were interviewed in Mandarin Chinese. All interviews were fully transcribed in the original interview languages for later analysis.

FINDINGS

Observed courses that were influenced by the GEPT

Angela's GEPT-Preparation course

Information from my whole semester of classroom observations as well as interviews with Angela and two of her students proved that Angela taught to the GEPT. Observations showed that her course was tailored, in terms of her teaching content, the homework she assigned, the tests she administered, and her teacher talk, to equip students with specific English abilities and test-taking strategies to take the GEPT. The teaching content and the assigned homework were adopted from commercial GEPT magazines. Quizzes, as well as mid-term and final examinations, were relevant to the GEPT because they were clone-like GEPT replicas in their format and content. Angela's teacher talk pertained to the GEPT, as reflected in the four categories of theme that emerged in her lectures: (1) GEPT test-taking strategies, (2) GEPT-pertinent information, (3) the school's resources and awards, and (4) censuring and prodding of students. In what follows, I will elaborate on each category.

Firstly, Angela proposed assorted test-taking strategies to tackle questions in different tested skills. For example, the following is one of her tactics to deal with the reading comprehension section:

You peek at what the questions ask first, and then go to the article to find answers. This will be faster and save time. Don't say that you don't have enough time and you

can't finish {answering all questions}⁸. What does this mean? You don't have a strategy to answer questions. [...] When there's a question that you don't know how to answer, do you want to think about it for a long time? No, it's a waste of your time and you won't be able to finish the rest of the questions. You can't answer this question, OK, skip to the next question. [...] You're uncertain about the eight questions, then circle and skip them. Later on when you finish all questions, [??]¹⁰ then revisit these questions, right? You can have time to <figure out,>¹¹ to guess.

This excerpt shows that Angela emphasized test-taking strategies. In the reading comprehension section of the GEPT, students needed to scurry to answer a multitude of questions under timed conditions; otherwise, they would leave questions unanswered, and the time for the test would be gone. In responding to the testing context, Angela proposed two strategies. Firstly, she advised students to peek at questions first and then return to the article to look for answers. Secondly, she suggested that they mark and skip tricky questions that they struggled with and revisit them later.

In addition to test-taking strategies, Angela also offered GEPT-relevant information in her classes, such as the registration periods. For example, she said in one of her classes:

Today is the last day for registering for the intermediate level of the exam. The registration form [????]¹² Test specifications seem to be available in our department.

This excerpt shows that Angela reminded students of the GEPT's registration period. In fact, she informed students of each registration period for the elementary and intermediate levels, which students were more likely to take. She also told students that the registration forms were available in the department for students' convenience. If she had not disseminated the information, students might have missed the registration deadlines.

Besides the GEPT information, Angela offered snippets of information in relation to awards, resources and assistance the department provided and encouraged students to capitalize on them. One case in point was that she urged students to use the self-study centre. Another example was that she disseminated the information that the department would earmark funds as incentives to students who passed the GEPT shortly after the plan was ratified by the school authority. She announced the following information in her class:

The GEPT motivation plan has been announced, right? [...] You can apply if you pass the first stage of the intermediate level. How much? //¹³ 3000 dollars¹⁴. How about passing the intermediate level? // 5000 dollars. Then you take the first stage of the high intermediate level? // 10,000 dollars. Oh, it's really worth it, right? [...] So, hurry! Take

⁸ { } : Words in the bracket represent my explanation of the interview or a situation.

⁹ [...] : Omission in the same utterance.

¹⁰ [??] : Inaudible for less than three seconds.

¹¹ < > : Code switching from Mandarin Chinese to English.

¹² [????] : Inaudible for more than three seconds.

¹³ // : A short pause in the teacher's lecture to allow students to answer the question.

¹⁴ One US dollar was worth approximately 33 new Taiwanese (NT) dollars at the time of writing.

this opportunity to register for taking the high-intermediate level. If you feel your English proficiency is good and you've passed the intermediate level, then take the first stage of the high-intermediate level. The first stage shouldn't be so difficult. [...] If you're in your fourth year of study, it should be a piece of cake for you.

The monetary incentive for students who had just passed the GEPT was a windfall. The incentive plan applied to students who would succeed in the GEPT in the future and who had passed the GEPT within six months prior to the announcement of the motivation plan. In fact, pecuniary awards were the most popular incentives among students according to my interviews with them.

The last category relevant to the GEPT in Angela's teacher talk was what I call the tactic of censuring and prodding. During my observation period, Angela berated some students several times for not paying attention to her lecture. Once she was irked by a student playing with a mobile phone, so she berated her: "You are gonna take an important test very soon. I don't know what you're doing." According to my classroom observations, a small portion of students frequently had whispered conversations with their adjacent classmates. In addition, a handful of students were sometimes absent, late for class or unable to complete their homework. Mostly, Angela would verbally encourage them or turn a blind eye. In a very few cases, Angela would censure these students as shown in this excerpt, reminding them of the imminent GEPT and their purpose for taking the GEPT-Preparation course.

I have described the impact of the GEPT on Angela's course, drawing on the evidence gathered from classroom observations. As the course title signified, I was not surprised to observe that her course was fashioned toward the GEPT. Her teaching content, testing content of, assigned homework and teacher talk left no doubt that her course was a GEPT-Preparation course.

My interview with Angela indicated that her deliberate practice was in accord with my interpretation that she taught directly to the GEPT, as shown in the following excerpt.

- Interviewer:* Were your teaching materials relevant to the GEPT?
Angela: The materials? Yes, they were all mock exam questions. Materials were all mock exam questions available in the market, or magazines, and so on.
- Interviewer:* Did you mention test-taking strategies in your class?
Angela: Yes! Yes! Sure! Sure! [...] I told them the effective ways to answer questions, so they even had time to check their answers.
- Interviewer:* Did you offer GEPT-relevant information in this course?
Angela: For example, I mentioned test dates. Then, I mentioned what they had to pay attention to. [...] Of course, I had to mention relevant information such as test items, trends, and current issues in the class. [...]
- Interviewer:* Did you encourage students to take the GEPT?
Angela: Of course I did. They had to.
- Interviewer:* Were quizzes, mid-term and final exams similar to the GEPT?
Angela: Yes.
- Interviewer:* Were assignments relevant to the GEPT?
Angela: Yes, for the reading comprehension section, they {students} had to try to answer questions in advance then came to class to discuss.
- Interviewer:* Overall, do you think you prepare students for the GEPT?

Angela: Yes.

Angela's testimony was mostly congruent with the results of my observations. As I observed, Angela also acknowledged that she used mock GEPT questions as teaching materials, coached test-taking strategies, offered GEPT-relevant information to students and encouraged students to take the GEPT. The other source of information about Angela's GEPT-Preparation course came from my interviews with her students, Summer and Jeffrey. Both of them reported that Angela taught to the GEPT. In conclusion, my classroom observations of and interviews with Angela and her students were consonant with one another, corroborating that her course prepared students for the GEPT.

Don's Listening and Speaking course

The results of my 16 hours of observation showed that the GEPT had an impact on Don's teaching content as well as mid-term and final examinations, but not on other aspects of his teaching. His teaching material was a monthly GEPT magazine that was available in local bookstores. Mid-term and final examinations were simulated GEPT examinations, which were produced by the same GEPT magazine publisher. On the other hand, Don never mentioned the GEPT explicitly in class, never offered GEPT-relevant information to his students, and did not instruct students in any test-taking strategies.

The results of my observations were mostly congruent with Don's testimony in his interview. He also believed that his teaching content was relevant to the GEPT, and the mid-term and final examinations were mock GEPT tests. However, he rarely coached students in test-taking skills and seldom offered students GEPT-relevant information.

Interviewer: Do you think the magazine you are using in this course is the material that prepares students for the GEPT?

Don: Yes, I think that's actually pretty good. [...] Because this one has vocabulary that students can see on their GEPT test, has English construction and grammar that they can see, it pushes their listening in ways that are similar to the GEPT test. [...]

Interviewer: Do you teach students test-taking strategies for the GEPT in class?

Don: Not so much in Listening and Speaking {course}. [...] [*****]¹⁵

Interviewer: Do you offer information that is relevant to the GEPT? [*****]

Don: Not so much because I really don't know that much about that stuff.

Interviewer: Do you encourage students to take the GEPT? [*****]

Don: Oh, Ya! Ya! Ya! They know. Because the rule here is that they must pass this level of the GEPT or its equivalent before they graduate. [...] So I encourage them and I tell them that actually they have no choice. [laughing] They must do it. [...]

Interviewer: Are test items and test types of the quizzes, mid-term exam, and final exam similar to those of the GEPT?

Don: The final exam and mid-term exam can be. [...]The quizzes, no. [...]

¹⁵ [*****]: Omission of one or several exchanges in the interview.

Interviewer: Are assignments relevant to GEPT preparation?

Don: [...] I think the assignments are helpful for that.

This excerpt provides data that are similar to those of my observations of Don's classes. His interview indicated that the teaching content was the easiest part of his teaching practice to be modified in response to the GEPT. Mid-term and final examinations were altered because the magazine publisher designed mock GEPT tests for consumers to use. Because the GEPT magazine author was Don's colleague, Don might adopt the magazine and its related mock GEPT tests out of convenience.

In comparison with Angela's course, the washback of the GEPT on Don's course was superficial. The teaching content and assessment could be altered immediately with limited effort. Except for these changes, no other washback was evident. One factor that had a bearing on this finding was his unfamiliarity with the test, as he acknowledged in his interview. As an expatriate teacher, his low Mandarin Chinese proficiency hindered his understanding of the GEPT, as well as the educational system in Taiwan. Another potential factor was his lack of affection toward teaching, which I felt from classroom observations. His lack of commitment was later corroborated by his resignation from his teaching post and his eventually taking a non-teaching position.

Like Don, four students from Don's course who participated the interviews also expressed their belief that Don's teaching content was tailored to prepare students for the GEPT, and his testing consisted of mock GEPT tests. To recapitulate, data I gathered from observations as well as interviews with Don and his students demonstrated a high degree of congruity, showing that the GEPT had an impact on Don's teaching content and testing.

Observed courses that were not influenced by the GEPT

The information elicited from my classroom observations, interviews with teachers and students was congruent, showing that the GEPT did not have any impact on both courses I observed at University A and on one observed course at University B. However, teachers' reasons for not teaching to the GEPT varied. The University A teacher, Carmella, did not teach to the test in her course of English Writing because she had negative views of the GEPT, and she objected to test-driven instruction.

Interviewer: Do you teach students test-taking strategies for the GEPT in class?
[*****]

Carmella: No, I don't because I don't have access to the questions on the GEPT. And my focus is the writing part and that's free writing. So and I also fundamentally disagree with the whole test-driven system that teaches to a test. I think that if you're learning English, you need to learn functional English, not test-taking.
[*****]

Interviewer: Do you encourage students to take the test after class?

Carmella: Well, I have problems with the GEPT because from looking at the questions the students have brought me, I think it's really barking up the wrong tree. It's asking very tricky, syntactic questions that have nothing to do with real English that any native speaker would go, what? What? What do you mean? The questions are almost like

puzzles [...] And I feel even the students can pass the GEPT, it does not prepare them to live in another country, or be proficient in English.

The other University A teacher, Sheila, said that she did not intend to teach to the GEPT when she designed her Listening and Speaking course. It seemed that the GEPT did not even occur to her when she planned and taught her course. In her interview, she did not consider the GEPT relevant to her course. Neither was the teaching of the University B teacher, Paul, influenced by the GEPT. In fact, in addition to his full-time teaching position at University B, Paul taught a GEPT-Preparation course part-time at a local cram school, so he was very cognizant with the GEPT. However, Paul said, "Teaching is one thing, and test preparation is another.... If I use test items to teach students, it'll not be like teaching at schools, but more like teaching at cram schools." His testimony suggested that he believed that formal schooling played a different role from cram schools; it had to focus on teaching and should not become test-driven. His belief in the roles formal schooling plays affected his teaching in his course of Reading and Vocabulary.

Courses that I did not observe at both universities

Other than the observed courses, I conducted interviews with 15 students at University A and 14 at University B. At University A, two teachers were said to administer mock GEPT tests in their courses for only two hours per semester, and another one was reported to offer extra-curricular GEPT lessons for her supervised students once or twice before they took the GEPT. According to students' accounts, the three teachers who administered the GEPT-relevant tasks did not spend much time on them. Their testimonies demonstrated that the GEPT had induced a limited impact on a handful of teachers at University A, although it was not adopted as a degree requirement, and teachers were not under pressure of losing their jobs or other undesired repercussions if they did not prepare students for the test. At University B, students said that three full-time teachers appeared to teach to the GEPT with commercial GEPT magazines. Their accounts meant that the teaching content had apparently been affected by the GEPT.

DISCUSSION

Although the GEPT had been universally taken by students in Taiwan and University B had even prescribed a GEPT requirement, this test induced some, but not a high level of washback on teaching in both applied foreign language departments I investigated. When I compare the GEPT's washback on both departments, there is not a striking difference except for the two courses taught by Angela and Don. This finding indicates that the GEPT requirement had a minor and teacher-specific impact on teaching practices.

Numerous factors affected the degree of washback in the present study. Overall, the objectives of the course and the relation of the course to the school's policy seemed to be the cardinal factors in determining the degree of washback on teaching. Other than these variables, some teacher factors need to be taken into account. For example, teachers' beliefs in the role of formal schooling, to some degree, determined the

washback of the GEPT. Moreover, teachers' perceptions of the test and their teaching philosophies affected their teaching. For example, Carmella had very negative views on the GEPT and was opposed to test-driven instruction, so she was reluctant to teach to the test. All of the above-mentioned arguments indicate that policymakers should consider teacher factors and micro-level contextual factors if they intend to use a test as a lever for change. The other important factor which has to be considered is that a high-stakes test for students may not bear equal significance for teachers. In this study, 17 out of 29 students considered the GEPT a high-stakes test, but teachers were not under pressure of losing employment or experiencing other undesirable repercussions if students performed poorly on the GEPT. This finding may also explain why the GEPT had a limited impact on teaching at both universities.

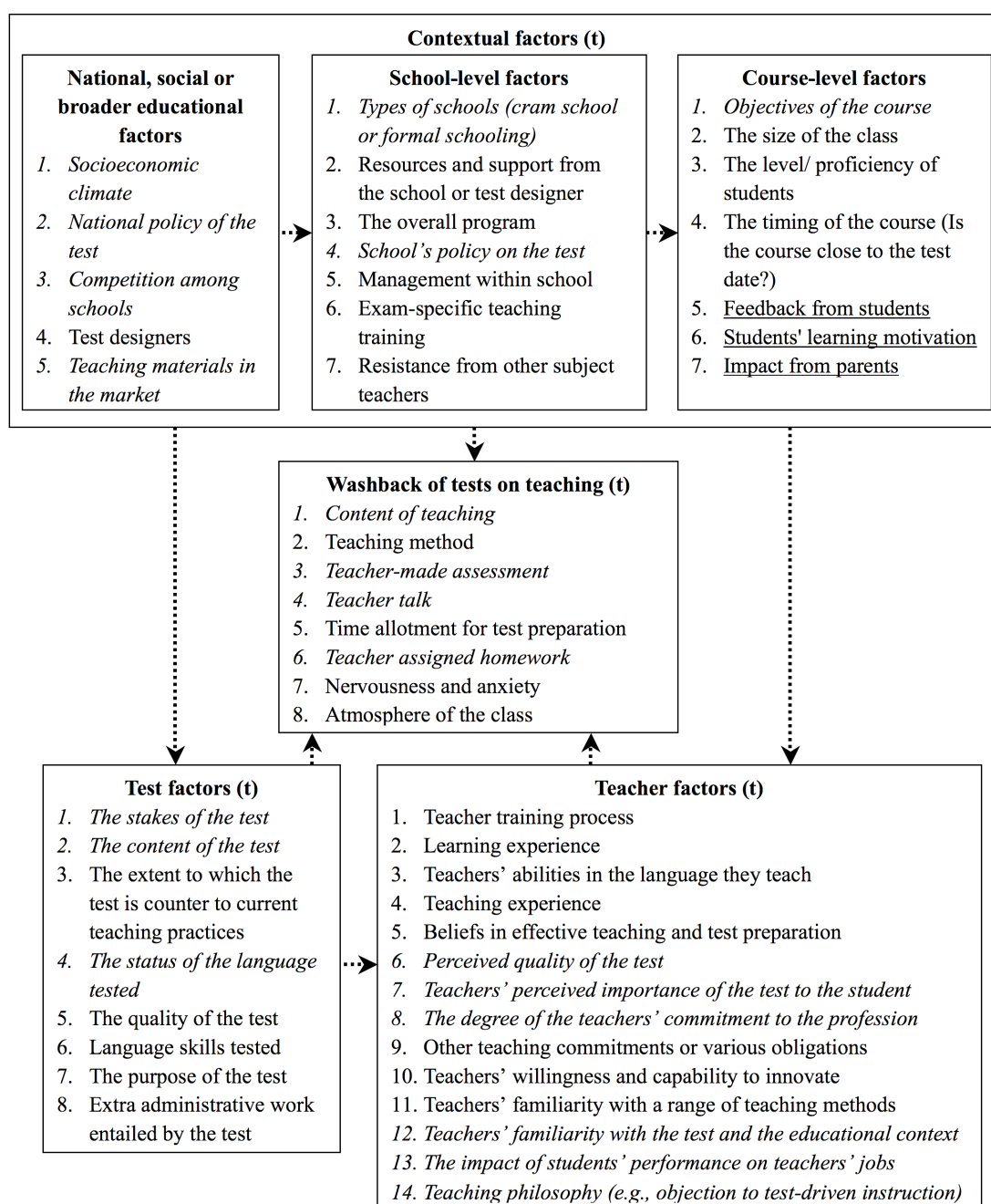


Figure 1. A model of washback.

As a researcher, I have often been asked how to predict or even promote positive washback. As shown in other empirical studies and my own research, washback varies from person to person and is a phenomenon which is inextricably linked to the contexts in which the test is administered. Therefore, manipulating and predicting washback is a daunting task. Factors that determine the washback of tests on teaching have been documented in different empirical studies, but no one has attempted to analyze and synthesize them in a figure or a table for reference purposes, an effort which I am convinced is beneficial when policy-makers, researchers and educators contemplate the potential washback tests can induce. Therefore, I propose a tentative model to schematise the washback of tests on teaching on the basis of current understandings of washback. By integrating all factors systematically on a single page, readers can infer potential washback effects more effectively.

In Figure 1, contextual factors, test factors and teacher factors influence the degree of washback on teaching. Dotted lines denote the impact of one category of factors on another. The symbol (t) acknowledges that washback phenomena may evolve over time, as Shohamy *et al.* (1996) pointed out. Factors in italics are either derived from this study or have been reported by other empirical studies, and are substantiated again in my study. Underlined factors have not been corroborated by any empirical data, but I believe that they are integral to understanding washback. The other factors, which have been elaborated on in the introductory section of the paper, have been documented in other empirical studies. Although enumerating all factors in each category, I do not mean that the figure contains all potential variables that play a role in the washback mechanism. Further research is still needed to deepen our understanding of washback.

Burrows' (2004) curriculum innovation model seems to focus on teachers' response patterns to the test. Complementing her model, mine mostly describes the factors that play a role in the washback mechanism, especially the contextual factors which are not mentioned in Burrows' model, but are indispensable to washback phenomena. In addition, this tentative model synthesizes factors reported in different empirical research studies and therefore can capture most facets of the washback mechanism. It also shows the interactions between different elements of washback, so readers can observe the dynamics of washback. To date, research has primarily focused on the top-down washback of tests on teaching or has looked at the impact of teachers' educational backgrounds or beliefs on their teaching. Little washback research has focused on how student factors play a role in affecting teaching within the washback mechanism. For example, how does students' feedback affect teaching? How does students' learning motivation influence teaching? Does parents' feedback have an impact on teaching? Do teachers' career plans, other teaching obligations, or family duties have an impact on their teaching within the washback domain? Further research is required to investigate these questions.

ACKNOWLEDGEMENTS

I thank Liying Cheng and Yoshinori Watanabe for their comments on the draft of this paper.

REFERENCES

- Alderson, J., & Banerjee, J. (2001). Language testing and assessment (Part 1). *Language Teaching*, 34, 213-236.
- Alderson, J., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13, 280-297.
- Alderson, J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Brennan, R. (2001). Some problems, pitfalls, and paradoxes in educational measurement. *Educational Measurement: Issues and Practice*, 20(4), 6-18.
- 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 (with 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, 38-54.
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes towards their English learning. *Studies in Educational Evaluation*, 24, 279-301.
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15, 253-271.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum.
- Chuang, W., & Hsu, Y. (2006). Washback effects of the General English Proficiency Test (GEPT) on English education in Taiwan. *Proceedings of the International Conference on English Language Teaching and Learning, Taiwan*, 298-310.
- Falk, B. (2002). Standards-based reforms: Problems and possibilities. *Phi Delta Kappan*, 83, 612-620.
- Ferman, I. (2004). The washback of an EFL National Oral Matriculation Test to teaching and learning. In L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods* (pp. 191-210). Mahwah, NJ: Lawrence Erlbaum.
- Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly*, 3, 333-368.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Haladyna, T., Nolen, S., & Haas, N. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge: Cambridge University Press.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods*

- (pp. 97-111). Mahwah, NJ: Lawrence Erlbaum.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Janesick, V. (2001). *The assessment debate: A reference handbook*. Santa Barbara, CA: ABC-CLIO.
- Jones, M., Jones, B., & Hargrove, T. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman & Littlefield.
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Lai, Y. (2003). Washback effects of the General English Proficiency Test (GEPT) on English language teaching and learning. *Proceedings of the 2003 National Conference on Linguistics*, Taiwan, 107-136.
- Language Training and Testing Center. (n.d.-a). *General English Proficiency Test*. Retrieved September 8, 2008, from http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_main.htm.
- Language Training and Testing Center. (n.d.-b). *Elementary*. Retrieved September 9, 2008 from http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_e.htm.
- Language Training and Testing Center. (n.d.-c). *Intermediate*. Retrieved September 9, 2008 from http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_i.htm.
- Language Training and Testing Center. (n.d.-d). *High-intermediate*. Retrieved September 9, 2008 from http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_hi.htm.
- Language Training and Testing Center. (n.d.-e). *Advanced*. Retrieved September 9, 2008 from http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_ad.htm.
- Ministry of Education. (2005, March 9). 教育施政主軸行動方案 [Four-year plan for 2005-2008 in education]. Retrieved April 19, 2005 from http://www.edu.tw/EDU_WEB/EDU_MGT/SECRETARY/EDU8354001/940309-01.doc.
- 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.
- Rich, W. (2003). Historical high-stakes policies relating to unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 22(1), 33-35.
- Shih, C. (2007). A new washback model of students' learning. *Canadian Modern Language Review*, 64, 135-162.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. Washington, DC: The National Foreign Language Center.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13, 299-317.
- Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington State. In L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods* (pp. 53-71). Mahwah, NJ: Lawrence Erlbaum.
- Vongpumivitch, V. (2006, June). An impact study of Taiwan's General English Proficiency Test (GEPT). Paper presented at the Language Testing Research Colloquium, Melbourne, Australia.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 291-302). Dordrecht, Netherlands: Kluwer.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A*

- case study using insights from testing and innovation theory*. Cambridge: Cambridge University Press.
- Wall, D., & Alderson, J. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10, 41-69.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 318- 333.
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129-146). Mahwah, NJ: Lawrence Erlbaum.
- Wu, R., & Chin, J. (2006). An impact study of the Intermediate-Level GEPT. *Proceedings of the 9th Academic Forum on English Language Testing in Asia*, Taiwan, 41-65.

Manuscript received: April 6, 2009

Revision received: May 31, 2009

Accepted: June 2, 2009

Appendix A: Framework for Field Notes for Classroom Observation

Date: _____

Course information			
Name of school		Division	
Grade & class		Title of the course	
Teacher's name		Number of students	
Starts at		Ends at	
Week no. (1-8)			

GEPT-related themes in teaching		
	1st Period	2nd Period
Test-taking strategies		
GEPT-relevant information		
Other verbal themes		
Test preparation materials		
Mock GEPT quizzes and exams		

GEPT-related themes in learning		
	1st Period	2nd Period
Students ask teachers GEPT-related questions		
Students discuss the GEPT in groups		

General description in learning		
	1st Period	2nd Period
Description of students' learning behaviour		

General description in teaching		
	1st Period	2nd Period
Description of teaching material		
Description of teaching method		

Comments and reflections:
Description of students' background:
Description of the classroom:

Appendix B: Interview Schedule for Teachers**The quality of the GEPT**

- ✓ What do you think about the advantages of the GEPT?
- ✓ What do you think about the disadvantages of the GEPT?
- ✓ Do you think that the GEPT is a valid test? Why?
- ✓ Do you think that the GEPT is a reliable test? Why?

Students in general

- ✓ What do you think about the learning motivation of students in your department?
- ✓ What do you think about the English ability of students in your department?

GEPT and teaching

- ✓ Do you think that the teaching materials you are using in your course are materials which prepare students for the GEPT?
- ✓ Do you teach students test-taking strategies for the GEPT in class?
- ✓ Do you offer students information relevant to the GEPT in class?
- ✓ Do you encourage students to take the GEPT in class?
- ✓ Are test items and item types (format) of quizzes, mid-term exam, and final exam similar to those of the GEPT?
- ✓ Are the assignments relevant to GEPT preparation?
- ✓ Overall, do you think that you prepare students for the GEPT in this course?
- ✓ Do you think that the course you are offering is helpful for students to take the GEPT?
- ✓ Have you ever thought of preparing students in this course? For what reason did you decide to prepare (or not prepare) students for the GEPT?
- ✓ Did you receive any support/training from the department for teaching GEPT-related courses or preparing students for the GEPT? If yes, what kind of support/training did you receive from the department?
- ✓ Has any student ever asked you to teach for the GEPT? Did you make changes on the basis of the student's request?