

英文畢業門檻政策對大學英文教師之影響調查

An investigation of the Impact of the English Benchmark Policy for Graduation on College EFL Teachers

朱秀瑜

Hsiu-yu Chu

摘要

本研究調查了英文畢業門檻政策對台灣兩所科技大學英文教師的影響。首先，研究者製作一份問卷，問卷的內容主要針對此影響的五個層面發問。接著，將製作完成的問卷發送給這兩所學校的英文教師填答。兩校總共有 35 位老師完成問卷。調查的結果顯示，英文畢業門檻政策對老師們的影響不大，且可能只停留在「表面」的層次。除此之外，老師所感受到有關英文畢業門檻考試所帶來的「利害關係」相當低，有可能是此政策對教師教學僅有微弱回沖效應的主要原因。

關鍵詞：英文畢業門檻政策、回沖效應、利害關係。

ABSTRACT

This present study investigated the impact of the English benchmark policy for graduation on teachers at two technological universities in Taiwan. A questionnaire on five aspects of the impact was developed and then conducted on teachers at both schools. A total of 35 teachers completed the questionnaire survey. The results show that the impact of the benchmark policy on teachers was not high and could be “superficial.” Furthermore, teachers’ perceived stakes of the English exit exam were relatively low, which might have led to the weak washback on their teaching.

Keywords : English benchmark policy for graduation, washback, stakes.

1. Introduction

The MOE of Taiwan has been trying to boost college students’ English proficiency by implementing a benchmark policy according to its 2005-2008 *Administration Guidelines*. As a result, more and more universities and colleges have implemented standardized English exit exams (such as the GEPT and its equivalents) on their students ever since. Unfortunately, even though it is still unknown whether the English proficiency benchmark policy has brought about any meaningful educational changes, there has been a striking discrepancy between the MOE’s expectation and students’ actual performance over the years, with the latter being much poorer than the former. Inevitably, almost every university and college which has set its English benchmark for graduation has had to provide other

options for those students who can not meet the requirement before graduation. The most common way is that students who are unable to pass the English exit exam will have to take extra courses to fulfill the graduation benchmark requirement instead. Some schools offer an internal test for students to choose to take on campus in addition to the external tests. A few others lowered the passing scores originally set for the GEPT test. In general, most universities and colleges have used external tests as their exit exams, but the supporting measures have varied from school to school.

It is actually not hard to find in the previous literature that the above-mentioned problem seems to have resulted from the tension between assessment for pedagogic purposes and assessment for administrative purposes, with the latter usually

winning out (McCay, 2006; cited in Wu, 2007). Though there might be some agreement between the two forces, more disagreement seems to have pulled the two forces apart. If the washback effects intended by the administration are to be achieved, those who serve the pedagogic purposes (the teachers) and those who are pedagogically involved (the students) will have to give in to the administrative force and move towards it. However, exactly how much further the pedagogic force will be pulled towards or away from the administrative force still remains a question for the present study. Figure 1 shows the tug-of-war between the two forces as a result of the benchmark policy. The overlap between a solid circle and any dotted circle represents the agreement between the two forces. The overlap could range from a large portion to none, suggesting that the pedagogic force may yield to or be entirely ungoverned by the administrative force.

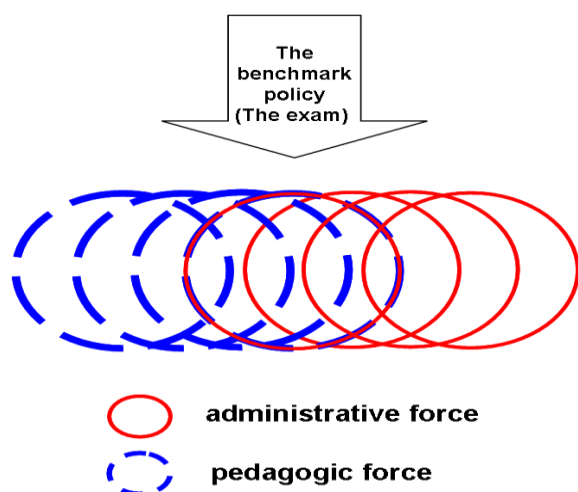


Figure 1. Tension between the pedagogic force and the administrative force as a result of the benchmark policy

As previously suggested, the pedagogic force in the above figure comes mainly from the teachers. This present study therefore focused on how much further the pedagogic force from the teachers was pulled towards or away from the administrative force. In other words, it is the impact of the benchmark policy on college EFL teachers that was investigated in this research project.

2. Literature Review

2.1 Washback on teaching content and teaching methods

Studies on the different aspects of teaching affected by the washback effects have come to different conclusions. Very often a contrast has been perceived between the teaching content and the teaching methods, with the former showing evidence of washback and the latter reflecting no sign of washback.

For example, Wall and Alderson (1993) in their Sri Lankan Impact Study claim that it seems that the exam in the research context affected only a small part of teachers on how they chose the teaching content but affected no teachers on how they taught in the classroom. Similar results are also found in Cheng's (1995) study on the influence of the introduction of an integrated and task-based approach into the existing Hong Kong Certificate of Education Examination in English (HKCEE). Cheng finds that textbooks played a very important role in the teaching of English in Hong Kong secondary schools. That is, when teaching the new syllabus, teachers did adopt different types of activities from the ones they had used before the introduction of the new exam; however, the obvious changes made in teaching lay in the different activities designed in the textbooks, which had come from the textbooks publishers' understanding of the new HKCEE. In this sense, the washback effect of the HKCEE on teachers' teaching in Hong Kong secondary schools was "superficial" because teachers seemed to teach what came next in the textbook.

However, as Spratt (2005) points out, the perception that washback affects teaching content but not teaching methods is not fully supported by the findings of some studies, such as Alderson and Hamp-Lyons (1996) and Watanabe (1996). The results of these two studies show that "whether the exam affects methods or not may also depend on factors other than the exam itself, such as the individual teacher" (Spratt, 2005, p.16).

In Alderson and Hamp-Lyons' study (1996), two teachers were observed teaching both the TOEFL preparation course and the non-TOEFL course. It was found that the two teachers used different teaching techniques in teaching the same course. However, their teaching techniques did not show much difference when comparing their own teaching in both the TOEFL preparation course and the non-TOEFL course. As a result, the researchers conclude that it is the individual teacher styles and personalities, instead of the TOEFL test, that clearly contribute to the nature of the classes observed.

Watanabe (1996) observes how exams influenced two teachers' teaching methods in a "yobiko" (preparatory school that offers various courses targeting specific university exams in Japan). He found that exams did have washback effects on one teacher's teaching methodology but not on the other's. The possible reasons for this finding, as proposed by Watanabe, might include (1) teachers' education background and experiences; (2) teachers' different beliefs about effective teaching methods and (3) the time related to the exam dates when the observations were made.

According to the above studies, teacher factors, rather than the test itself, play an important role in affecting the washback effects on teaching methodology.

2.2 Washback on teachers' attitudes and feelings

The impact of tests on teachers' attitudes and feelings towards teaching and testing is another important issue that has frequently appeared in the previous literature. Results from the previous studies are mixed. Some teachers showed negative attitudes and feelings towards the exam because they felt pressured teaching towards the test (Alderson and Hamp Lyons, 1996; Shohamy et al., 1996), or because they are worried about not being able to improve their students' scores (Alderson and Hamp Lyons, 1996) or not being able to make the test-preparation class interesting (Alderson and Hamp

Lyons, 1996; Read and Hayes, 2003). However, some teachers showed positive attitudes and feelings towards the exam because they believed the new test would push the teachers to teach communicative skills (Cheng, 1999; Shohamy et al., 1996) or because they felt they were helping students cope with something important (Alderson and Hamp-Lyons, 1996). In terms of the GEPT test, Wu and Chin (2006) find that most of the teachers participating in the interviews and questionnaire survey reacted positively to the implementation of the four-skill test, perceiving the importance of listening and speaking in the curriculum. However, the classroom observation data show that teachers spent little time preparing students for the test, because most of the class time was still devoted to the two higher-stakes university entrance exams, which test only reading and writing.

3. Methodology

3.1 Research Context

The present study was situated in two technological universities in Taiwan. These two universities, marked as School A and School B, had quite similar student backgrounds and English graduation benchmark policies. School A, formerly a private five-year junior college, was upgraded into a technological university which recruited mostly vocational high school graduates for its four-year undergraduate program, with an increasing enrollment of high school graduates in recent years. There were approximately 2500 students majoring in engineering-related subjects in three colleges. School B was a public technological university, which used to be a five-year junior college as well. The majority of the students in this school majored in engineering-related subjects like School A. The total number of students in School B was around 8000.

In addition to the four-year undergraduate program, both schools also had a two-year undergraduate program which recruited five-year junior college graduates. However, the two-year

undergraduate program would sooner or later become history, because five-year junior colleges would no longer exist in Taiwan's higher education and there was no need keeping a two-year undergraduate program. Students graduated from either the four-year or two-year undergraduate program are awarded a bachelor's degree.

Both schools had similar graduation benchmarks for English. They both required their students to pass the first stage of the GEPT Intermediate test (or other equivalent tests). However, there were some differences between the two schools' make-up measures. In School A, freshman students in the four-year undergraduate program were required to take an internal test (equivalent to the first stage of the GEPT Intermediate test), serving as an exit test and held twice a year, if they hadn't passed the graduation benchmark by the time the test was held. Sophomores, juniors and seniors in School A might choose to take the test with the freshmen. In School B, all the freshman and sophomore students in the four-year undergraduate program and all the junior students in the two-year undergraduate program had to take an English proficiency test (each covering five pre-scheduled GEPT mock tests on the school's website) right after their mid-term and final exam. Students' scores on the two English proficiency tests accounted for 30% of the overall grades for the required English course in that semester.

Students in both schools had to take a one-semester make-up course to fulfill the graduation benchmark requirement if they still had not passed the English exit exam later in their college life. In school A, the make-up course was offered in the first and second semester of the senior year for the four-year undergraduate program students; in school B, the make-up course was offered in the second semester of the junior year for the four-year undergraduate program students and in the first semester of the junior year for the two-year undergraduate program students. In addition, for students to take the make-up course in School B, the

prerequisite was that they must have already taken the external GEPT test (the first stage of the Intermediate test) and gained a total score no less than 80 points on listening and reading. To pass the make-up course in School B, students also had to take the English proficiency test (as mentioned earlier) after the mid-term and final exam and must gain from the two tests an average score of over 60 points on the listening and reading component respectively. In both schools, if students did not pass the make-up course, they would have to re-take the course until they pass it.

School A had implemented the policy on only its four-year undergraduate program students since the 2004 academic year, while School B on both its four-year and two-year undergraduate program students since the 2003 academic year, with continuing changes over the years to the above-mentioned latest version.

3.2 Instrument and Participants

The main instrument used in the present study was the Teacher's Questionnaire (TQ). The development of the TQ will be illustrated in the following section, as well as the background information of the teachers completing the TQ.

3.2.1 The Teacher's Questionnaire (TQ)

The preliminary TQ, consisting of 37 items in total, was divided into three parts. The first part was to elicit the participant's background information. The second part of the TQ was to collect information on the washback effects of the English exit exam on teachers, falling into three dimensions. The first dimension of the impact looked at teachers' attitudes towards the exam and the exam-related issues. The second dimension of the impact focused on what teachers had been teaching, centering around their syllabi and teaching materials. The third dimension of the impact probed into how teachers had been teaching, concentrating on their classroom activities and assessment. Finally, the third part of the TQ was to investigate teachers' perceived consequences of the

graduation benchmark policy for English.

A panel of five experts was invited to establish the content validity and face validity of the TQ. These five experts included two university professors specializing in TESOL theories and instrumentation, and three professors of Applied English at two technological universities other than School A and B, which had also implemented a graduation benchmark policy.

The English and Chinese versions of the preliminary TQ, as well as a cover letter and two accompanying questionnaire item validation forms, were first sent to the panel of the five experts. The panel was asked to review each item of the two questionnaires for its appropriateness and clarity and to indicate on the questionnaire item validation forms whether the item was appropriate and clear. For any item judged as appropriate but unclear, the panel was asked to provide a rephrased statement to improve its clarity. For any item judged as inappropriate by two or more experts on the panel, it was deleted. All the items were revised based on these criteria, and the number of items was thus reduced from 37 to 29 for the TQ. All variables were assessed using a 6-point Likert scale. The revised version of the TQ was then ready for factor analysis and the reliability test.

The revised Chinese version of the TQ was pilot tested on 65 teachers. These teachers were from schools other than School A or School B but had similar backgrounds.

The principal-components analysis and factor analysis (Varimax rotation) on the revised TQ items identified five factors that constitute teachers' perceived washback, status and stakes of the English exit exam. The first three factors, renamed as (1) teachers' worries about their students' performance; (2) teachers' indifference to the English exit exam and (3) teachers' adaptive teaching to the English exit exam, represented the three dimensions of the washback effects of the English exit exam on teachers. The last two factors, renamed as (4)

teachers' perceived stakes of the English exit exam and (5) teachers' perceived status of the English exit exam, accommodated the same items as the preliminary TQ.

3.2.2 *Teacher participants*

A total of 35 teachers completed and returned their questionnaires. The background information about these 35 teachers is briefly summarized as follows: 14 of them were from School A and 21 from School B. Teachers from both schools shared quite similar features in every respect except that all the School A teachers were lecturers holding a master's degree, while School B teachers varied in their teaching positions (4 associate professors, 4 assistant professors and 13 lecturers) and highest degrees (7 Ph.D.s, 13 M.A.s and 1 B.A.).

3.3 Data Collection and Analysis

Data were collected in the spring semester, 2008. In the following sections, data collection and data analysis procedures will be presented in details.

3.3.1 *Data Collection Procedures*

Since the number of teachers at both schools was limited, the researcher delivered the TQ in person to most of the full-time and part-time teachers available at both schools. In order to guarantee a higher response rate, the teachers were told that the researcher would come back to collect the completed questionnaire later the day or one week later. The response rate was thus greatly increased in this way, with 35 teachers returning their completed questionnaires, as mentioned in the previous section.

3.3.2 *Scoring of the Teacher's Questionnaire (TQ)*

To decide on the extent to which the washback effects of the English exit exam were on teachers in each of the three dimensions (Factor 1, 2 and 3), three mean scores, each for one dimension, were obtained.

Teachers' perceived consequences of the graduation benchmark policy were scored as follows: a mean score obtained for Factor 4 measured the degree of teachers' perceived stakes (immediate

importance) of the English exit exam. Another mean score was obtained for Factor 5 to show teachers' perceived status of the English exit exam.

4. Results and Discussion

4.1 The three dimensions of teachers' perceived impact of the English exit exam

Table 1 shows the descriptive statistics of the three dimensions of teachers' perceived impact of the English exit exam (*teachers' worries about their students' performance, teachers' indifference to the English exit exam and teachers' adaptive teaching to the English exit exam*). The statistics were transformed into a bar chart in Figure 2.

Table 1. Descriptive statistics for teachers' perceived impact of the English exit exam on the three dimensions

	N	Minimum	Maximum	Mean	SD
Worries	35	1.00	6.00	3.88	1.15
Indifference	35	1.33	5.33	3.49	.89
Adaptive teaching	35	1.40	5.30	3.75	.91

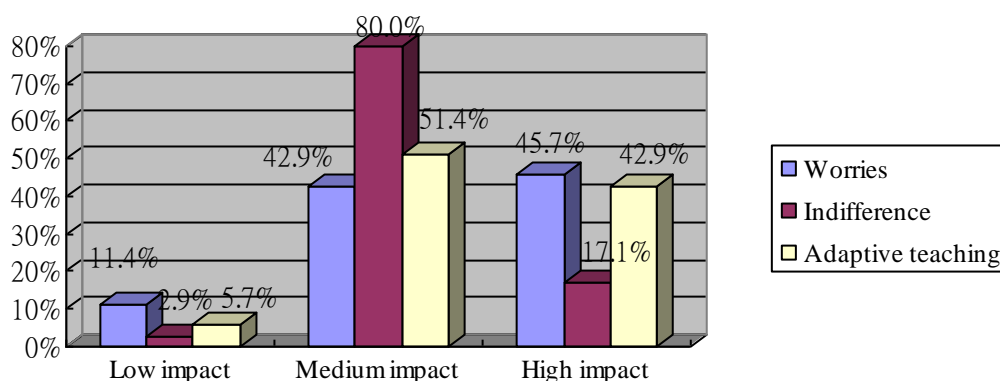


Figure 2. Distribution of teachers' perceived impact of the English exit exam on the three dimensions

The descriptive statistics in Table 1 show that the mean score for *teachers' worries about their students' performance* (3.88) is slightly higher than the other two dimensions (3.49 and 3.75). In Figure 2, teachers were reorganized into three groups (low, medium and high) in terms of each dimension. For those who obtained an average score lower than 2.00, they were categorized into the low impact group; 2.01—4.00 into the medium impact group and higher than 4.00 into the high impact group. The majority of the teacher participants perceived medium to high impact on the three dimensions. It is worth noticing that more teachers fall into the high impact group

than the medium impact group when it comes to their worries about students. A huge number of teachers (80.0%) show a medium degree of indifference to the English exit exam.

However, there is no significant difference among the three dimensions of impact as tested by the repeated-measures ANOVA, indicating that teachers did not perceive the three dimensions of impact differently. The results are summarized in Table 2. A closer examination of the TQ items was conducted by looking at the mean score for each item, as shown in Table 3.

Table 2. Repeated-measures ANOVA on the three dimensions of teachers' perceived impact of the English exit exam

Source	SS	df	MS	F	p
Impact dimensions	2.77	2	1.38	1.28	.28
Subjects	27.12	34	.80		
Residual	73.28	68	1.08		
Total	103.17	104			

Table 3. A list of mean scores for the TQ items on the three dimensions of teachers' perceived impact

No.	Item	Mean
Teachers' worries about their students' performance		(1-6)
10	I feel frustrated at my students' low chances of passing the English exit exam due to their low English proficiency.	3.69
9	I feel pressured by my students' performance on the English exit exam.	3.86
11	I am worried that my students would fail in the English exit exam.	4.09
Teachers' indifference to the English exit exam		(1-6)
18	The English exit exam has little impact on what I teach.	3.63
24	The English exit exam has little impact on how I teach.	3.51
13	I pay little attention to the English exit exam while constructing my teaching syllabus.	3.31
Teachers' adaptive teaching to the English exit exam		(1-6)
19	I arrange my classroom activities to meet the requirements for the English exit exam.	3.71
16	The English exit exam influences what supplementary materials I use.	3.86
14	I include relevant materials in my teaching syllabus to cover the possible subject matters on the English exit exam.	4.14
21	I adapt items from the mock tests of the GEPT or other equivalent language tests as quizzes to test my students.	3.63
20	I teach test-taking strategies, especially as the English exit exam date gets closer.	3.74
17	I need to prepare more related materials for students because of the English exit exam.	3.80
15	I include mock tests of the GEPT or other equivalent language tests in my teaching materials.	4.00
23	I adjust my class time allotment of different kinds of activities due to the English exit exam.	3.29
12	I am willing to teach towards the English exit exam.	4.09
22	I adjust my assessment on students' learning due to the English exit exam.	3.20

For *teachers' worries about their students' performance*, teachers show a mean score high at 4.09 on a six-point scale for item (11) (*I am worried that my students would fail in the English exit exam*). For *teachers' indifference to the English exit exam*, all the three items show a similarly medium degree of teachers' indifference to the English exit exam. For *teachers' adaptive teaching to the English exit exam*, the mean scores for items (15), (20) and (21) on the use of mock tests and teaching of test-taking strategies are between 3.63 and 4.00, and the mean scores for items (17), (16) and (14) on teaching

materials range from 3.80 to 4.14, both of which are a little bit higher than the medium, suggesting the use of the GEPT mock tests might have played a part in terms of teachers' teaching materials change. The mean scores for items (19) and (23) on teaching activities reach 3.71 and 3.29 respectively on a six-point scale, suggesting there might not be a strong impact on that part.

The most interesting part on Table 3 is about items (14), (12) and (13). Items (14) and (12) got the top two scores (4.14 and 4.09) among all items,

suggesting that it seems to be fine with quite a few teachers to use test-related materials and to teach towards the English exit exam. However, the medium score (3.31) for item (12) seems to imply that not many teachers ever thought about planning their teaching towards the English exit exam (also supported by the medium scores for items (18) and (24)). It may be inferred from the conflicting results here that most teachers still stuck to their own teaching, which might not be directly related to the test, but were willing to add test-related materials, such as mock tests, to their teaching. However, these test-related materials and teaching were simply supplementary and extra to their main teaching plans.

In other words, the change of teachers' teaching caused by the English exit exam, if any, was only "superficial."

4.2 Teachers' perceived stakes of the English exit exam

As shown in Table 4, the mean score for teachers' perceived stakes of the English exit exam is fairly low at 2.37 on a six-point scale, much lower than the *status* perceived by teachers (4.37). The difference between teachers' perceived *stakes* and *status* was significant using the paired-samples *t*-test, as shown in Table 5.

Table 4. Descriptive statistics for teachers' perceived stakes and status of the English exit exam

	N	Minimum	Maximum	Mean	SD
Stakes	35	1.00	4.50	2.37	.97
Status	35	2.67	6.00	4.37	.78

Table 5. Paired-samples *t*-test on teachers' perceived stakes and status of the English exit exam

	<i>df</i>	<i>t</i>	<i>p</i>
Stakes vs. Status	34	-10.99*	.00

* $p < .05$

Previous literature suggests that washback effects on learning are closely related to the stakes of the test (Shohamy, 1993), while the status of the test does not actually influence students in terms of washback (Stoneman, 2005). In a similar vein, the washback of the GEPT test on teaching in the present study was low possibly due to its low stakes perceived by the teachers at both schools.

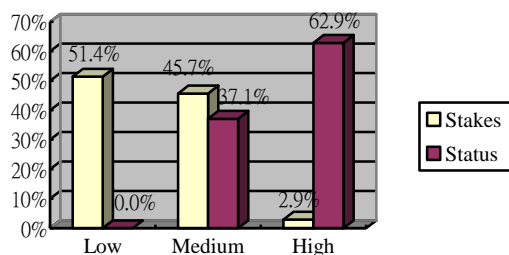


Figure 3. Teachers' perceived stakes and status of the English exit exam

Figure 3 shows the comparison among the low, medium and high groups in terms of teachers' perceived stakes and status. The three groups were divided according to their average scores on each issue (low group: lower than 2.00, medium group: 2.01—4.00, high group: higher than 4.00). It is apparent that the majority of teachers perceived low (51.4%) and medium (45.7%) stakes of the English exit exam, while a large number of teachers (62.9%) perceived a high status of the English exit exam.

A closer look at the TQ items in Table 6 about the stakes issues shows that students' GEPT test performance had little influence on teachers' evaluation (2.63) and their teaching jobs (2.61). As for the status issues, teachers perceived a higher status of the internationally-recognized language tests (such as the TOEFL and TOEIC) than the locally-developed GEPT test (4.91 vs. 3.94), even

though the latter was the major test they had intentionally pushed their students to take. Teachers also associated the implementation of the graduation

benchmark policy with their school's reputation (4.26), suggesting the policy had been socially-justified.

Table 6. A list of mean scores for the TQ items on teachers' perceived stakes and status of the English exit exam

No.	Item	Mean
Teachers' perceived stakes of the English exit exam		(1-6)
28	I think students' external test (such as the GEPT or other equivalent language tests) results have a great influence on my evaluation in this school.	2.63
29	I think students' external test (such as the GEPT or other equivalent language tests) results have a great influence on my holding a teaching post in this school.	2.11
Teachers' perceived status of the English exit exam		(1-6)
27	I think the international language tests (such as the TOEFL, TOEIC and IELTS etc.), serving as the English exit exam for English in our school, are publicly-recognized tests. Their score reports could be used as a certificate of students' English proficiency for their pursuit of higher education or future job hunting.	4.91
26	I think the GEPT, serving as the English exit exam for English in our school, is a publicly-recognized test. Its score report could be used as a certificate of students' English proficiency for their pursuit of higher education or future job hunting.	3.94
25	I think the implementation of the English benchmark for graduation is very important to the reputation of our school.	4.26

In summary, teachers perceived low stakes but a much higher status of the English exit exam. In addition, teachers seemed to recognize the international language tests more than the local GEPT test, even though they were still pushing the GEPT for the graduation benchmark policy.

5. Conclusion

The TQ survey results show that the three dimensions of teachers' perceived impact of the English exit exam (*teachers' worries about their students' performance, teachers' indifference to the English exit exam and teachers' adaptive teaching to the English exit exam*) do not differ significantly, though *teachers' worries about their students' performance* has a slightly higher mean score than the other two dimensions. A closer look at the TQ items shows the following results: For *teachers' worries about their students' performance*, most teachers did worry about their students for not being able to pass the GEPT test. For *teachers' indifference to the English exit exam*, teachers showed a medium degree of indifference. For *teachers' adaptive*

teaching to the English exit exam, teachers showed a little bit higher degree of impact on their teaching materials and a medium degree of impact on their teaching activities. However, the impact on both teaching materials and activities might only be "superficial," as raised in Cheng (1995), since some teachers seemed to have taken the inclusion of the GEPT mock tests as a major change in their teaching.

One interesting finding from the TQ survey results is that it is not entirely impossible for teachers to teach to the English exit exam. However, it still remains a question under what circumstances will teachers start to teach to the test, because teachers in the present study did not show any more substantial changes than the above-mentioned superficial washback.

On the other hand, teachers' perceived stakes of the English exit exam were relatively low, indicating that no concrete consequences had been actually imposed on teachers due to the graduation benchmark policy. As suggested in Alderson and Wall's Washback Hypothesis (*Tests that have important*

consequences will have washback; and conversely. Tests that do not have important consequences will have no washback), teachers' perceived low stakes of the English exit exam might have led to the weak washback on teaching. It might thus be concluded that the pedagogic force, as a result of the benchmark policy shown in Figure 1, might not have yielded to the administrative force. In other words, the impact of the benchmark policy might not have been as high as some had expected.

References

1. Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics* 14 (2), 115-129.
2. Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing* 13 (3), 280-297.
3. Cheng, L. (1995). How does washback influence teaching? Implications for Hong Kong. *Language and Education* 11, 38-54.
4. Cheng, L. (1999). Changing assessment: washback on teacher perceptions and actions. *Teaching and Teacher Education* 15, 253-271.
5. Cheng, L. (2005). *Changing language teaching through language testing: A washback study. Studies in Language Testing: Vol. 21*. Cambridge: Cambridge University Press.
6. Cheng, L., & Watanabe, Y. (with A. Curtis) (Eds.) (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum.
7. McKay, P. (2006). *Assessing young language learners*. Cambridge Language Assessment Series. Cambridge: Cambridge University Press.
8. Ministry of Education. 2007 *Administration Guidelines*. (in Chinese) 教育部 96 年度施政計畫 http://www.edu.tw/files/site_content/B0039/教育部_96年度施政計畫.doc
9. Read, J. and Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. *IELTS International English Language Testing System Research Reports* 4, 153-206
10. Shohamy, E. (1993). The power of test: The impact of language tests on teaching and learning. NFLC Occasional Papers. ED 362-040
11. Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing* 13 (3), 298-317.
12. Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research* 9 (1), 5-29.
13. Stoneman, B. W. H. (2005). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviors*. Unpublished doctoral dissertation, The Hong Kong Polytechnic University, Hong Kong.
14. Watanabe, Y. (1996). Does grammar translation come form the entrance examination? Preliminary findings from classroom-based research. *Language Testing* 13 (3), 318-333.
15. Wu, J. R. W. (2007). English language assessment in Taiwan: Where do we go from here? *Proceedings of the 2007 International Conference and Workshop on TEFL & Applied Linguistics* (pp. 574-586). Taipei: Crane Publishing Company.
16. Wu, R. & Chin, J. (2006). An impact study of the Intermediate-level GEPT. *Proceedings of the Ninth Academic Forum on English Language Testing in Asia* (pp. 41-65). Taipei., Taiwan.