

# English Benchmark Policy for Graduation in Taiwan's Higher Education: Investigation and Reflection

Hsiu-yu Chu

General Education Center, Ming Chi University of Technology, Taipei, Taiwan

Hsi-nan Yeh

Department of English, National Taiwan Normal University, Taipei, Taiwan

**Abstract**—Encouraged by the Ministry of Education (MOE), an increasing number of universities and colleges in Taiwan have started to set their own English benchmarks for graduation using external standardized language tests as exit exams. The present study investigated and reflected on the implementation of this policy at two technological universities in Taiwan by answering two research questions: (1) How did college administrators, teachers and students perceive the policy? (2) What washback effects did the English graduation benchmark policy bring about? A mixed-methods approach was adopted for the study. The results show that although there was a social consensus about the implementation of the policy, the policy did not seem to have achieved what the MOE had expected due to its limited and weak washback on teaching and learning. It is suggested that the English graduation requirements might have to be modified in the subsequent policies for more meaningful and practical results.

**Index Terms**—English benchmark, policy, graduation, higher education, Taiwan

## I. INTRODUCTION

In many Asian countries, where examinations have long played an important role in their culture and history, decision-making bodies in education tend to view tests as a panacea for educational innovation. Chen et al. (2005) in a study on Taiwanese students' language learning motivation even added one more category, the *required orientation*, to the two existing categories, the integrative orientation and instrumental orientation, to signify the important role tests have played in most Taiwanese students' language learning experience. With this exam-oriented tradition, it was not surprising to see that the Ministry of Education (MOE) of Taiwan embodied an English proficiency benchmark policy for college undergraduates in its *2005-2008 Administration Guidelines*<sup>1</sup> as one of the strategies to promote Taiwan's globalization, as required by the government's *Challenge 2008: National Development Plan* (Council for Economic Planning and Development [CEPD] 2008).

*Challenge 2008: National Development Plan* featured ten key individual plans including "e-generation manpower cultivation plan," which aimed to cultivate e-generation manpower with good IT, English, and creative skills. Three measures were taken for this individual plan, one of which was to "create an internationalized daily environment and increase all people's English proficiency on the island." Five yardsticks were used to evaluate the results of this measure: (1) the rate of foreign visitors being satisfied with the English environment in Taiwan, such as bilingual signs, publications and websites; (2) college students' GEPT (General English Proficiency Test, a local test) passing rates at the elementary and intermediate levels; (3) civil servants' GEPT passing rates at the elementary level; (4) the number of foreign students studying for degrees in Taiwan; (5) the number of Taiwanese students studying abroad. It is not hard to see that, in addition to the first yardstick, the goal was practically test-oriented and school-based. Judging from Spolsky's (2004) three components of a language policy of a speech community, namely, its language practices, its language beliefs or ideology and any specific efforts to modify or influence that practice by any kind of language intervention, planning or management, Taiwan's English language policy was far from generating concrete linguistic results. First of all, in terms of language practices, the policy did not aim to make English one of its "linguistic repertoire" (Spolsky 2004, p.5), since the so-called "English environment in Taiwan" was limited only at a non-interactive level, such as bilingual displays, instead of using English as a daily communication medium. Second, most Taiwanese people's beliefs about English and its use still view English as a school subject (Simpson 2007) and a tool for better job opportunities, as reflected in the test-oriented and school-based implementation of the policy. Third, the efforts to modify or influence the attempted language practice were not strong enough to bring about any significant

<sup>1</sup> For more information about Taiwan MOE's *2005-2008 Administration Guidelines*, please refer to <http://english.moe.gov.tw/ct.asp?xItem=7043&ctNode=784&mp=1#1>.

changes. The present study will focus on the third component with specific examples from the college settings in Taiwan.

## II. BACKGROUND AND RATIONALE

### A. Taiwan's Higher Education System

Under the current education system in Taiwan, there are two types of undergraduate programs. One type is the four-year university undergraduate programs, which recruit mostly high school graduates. The other type constitutes technological colleges or universities in the Technological and Vocational Education (TVE) system, which offers (1) four-year undergraduate programs, mainly for vocational high school graduates, and (2) two-year undergraduate programs, particularly for five-year junior college (starting after the junior high school) graduates and two-year junior college (starting after the vocational high school) graduates. Students graduated from any of the above undergraduate programs are awarded a bachelors' degree.

### B. The MOE's Administration Guidelines

According to the MOE's 2005-2008 Administration Guidelines, the MOE encouraged each university and college to set their own English benchmark for graduation, using the GEPT or other language tests, such as the TOEFL, TOEIC and IELTS. The MOE's preferred threshold was CEFR-B1 for four-year university graduates and CEFR-A2 for those in the TVE system.

The GEPT test is a local test developed by the Language Training & Testing Center (LTTC) in Taiwan since 1999 and is administered at five levels: elementary, intermediate, high-Intermediate, advanced and superior. Each of the first four levels is administered at two stages: The first stage includes the listening and reading components and the second stage includes the speaking and writing components. Examinees must pass the first stage before proceeding on to the second. The superior level is an integrated test of all four skills. Each level approximates to CEFR A2 (elementary), B1 (intermediate), B2 (high-Intermediate), C1 (advanced) and C2 (superior) respectively (Wu and Wu 2007).

In general, most four-year universities and colleges set their graduation benchmark for non-English majors at the first stage of the high-intermediate or the second stage of the intermediate level of the GEPT test, while most technological universities and colleges set theirs at the first stage of the intermediate or the second stage of the elementary level. For English majors in both systems, the benchmark was usually higher than that for their non-English major peers.

### C. Problems with the English Benchmark Policy for Graduation

This trend reflected the strong demand of English in Taiwan's higher education and, what is more, the power of the MOE authority. As Shohamy (1993) put it, "using tests to solve educational problems is a simplistic approach to a complex problem. It works on people's fear of authority" (p.19). In her point of view, the power of tests is too strong to bring about any meaningful educational changes, since the changes are often instrumental.

Inevitably, almost every university and college which set its English benchmark for graduation had to provide other options for students who could not meet the requirement before graduation. The most common way was for those students to take extra courses to fulfill the graduation benchmark requirement instead. Some schools offered an internal test for students 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 used external tests (the GEPT and its equivalents) as their exit exams, but the supporting measures varied from school to school.

As pointed out by Wu (2007), a serious problem thus arose with these supporting measures for the graduation benchmark requirement; that is, the criteria set for each of the measures for the graduation benchmark requirement did not seem to align with one another. There was no evidence to show that a student passing an internal exam or an extra course had the same level of required English proficiency for graduation as those who passed the external GEPT test.

To sum up, not all college students were able to pass the required English exit exam. Instead, they had to take the internal English exit test or an extra course to fulfill the graduation requirements. However, there seemed to be an alignment problem between each individual school's English benchmark and their substitute measures. Two research questions were thus formulated for the present investigation.

1. How did college administrators, teachers and students perceive the English benchmark policy for graduation?
2. What washback effects did the policy bring about?

## III. METHODOLOGY

### A. Research Context

The present study was situated at 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. Most students at each school, where a four-year as well as a two-year undergraduate program was offered, were engineering majors. Both schools required their students to pass the first stage of the GEPT intermediate test (or other equivalent tests) as the graduation benchmark with slightly different 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 still hadn't passed the external test by the time the internal test was held. Sophomores, juniors and seniors in School A might choose to take the internal test with the freshmen if needed. In School B, all the freshman and sophomore students in the four-year undergraduate program and all the junior students<sup>2</sup> 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 exams. Students' scores on the two English proficiency tests accounted for 30% of the overall grades for the required English course in that semester.

Students at 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 college. In school A, the make-up course was offered in the first and second semesters of the senior year for the four-year undergraduate program students, while 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 second 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<sup>3</sup>. 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 exams and must gain from the two tests an average score of over 60 points on the listening and reading components respectively. At both schools, if students did not pass the make-up course, they would have to re-take the course until they pass it.

### B. Participants

A total of ten teachers, six administrators and 32 students at both schools participated in the interviews; eight teachers' classes were observed and their teaching and testing materials were analyzed. For these participants' background information, please refer to the Appendices. Meanwhile, 633 students agreed to take the questionnaire survey.

### C. Instruments

*Interview guides:* A semi-structured interview approach was adopted. All the interview questions for the two research questions were put together in one single interview guide for each group of stakeholders. The Chinese version of the interview guides was reviewed by two PhD students in TESOL to ensure their appropriateness and clarity. The revised interview guides were then tested on several similar participants, who were not included in the formal study, for their effectiveness. The interview guides were revised accordingly before they served as the research instrument for the formal study.

*Observation forms:* In addition to verifying what the teachers said in the interviews, several required courses and one make-up course were observed at each school in an attempt to find more themes that might not have emerged in other sources of data. An onlooker observation approach (Patton 2002) was adopted, using two observation forms, one for the required English classes and one for the make-up course. The observation forms were developed with major "sensitizing concepts" that serve as a guide to help manage the observational task (Patton 2002, p.279).

*Students' questionnaire (SQ):* To investigate the washback effects of the English graduation benchmark on students, a students' questionnaire (SQ) was developed based on Alderson and Wall's (1993) Washback Hypothesis and Spratt's (2005) categorization of areas of washback effects, then collapsed into three major dimensions: (1) impact on attitudes; (2) impact on 'what' and (3) impact on 'how.' The first dimension of the washback effects is involved with attitudinal changes, while the second and third dimensions are concerned about behavioral changes.

The principal-components analysis and factor analysis (Varimax rotation) on the SQ items identified three factors that constitute three dimensions of the washback of the English graduation benchmark on learning, renamed as (1) students' perceived impact on their out-of-school practice, (2) students' perceived impact on their efforts and (3) students' worries about the graduation benchmark, which represent the three dimensions of the washback effects of the policy on students.

### D. Data Collection

Teachers were recruited in person or by phone calls. In School A, two teachers teaching required courses agreed to be randomly observed for six and five weeks (two hours each week) respectively. Another teacher teaching the make-up course agreed to be observed once (two hours). In School B, four teachers agreed to be observed for two weeks and one teacher for only one week (two hours each week). Throughout the whole observation period, teaching and testing materials such as syllabi, textbooks, class handouts, teacher-made tests and students' homework were collected whenever possible.

<sup>2</sup> Students in the first and second year of the two-year undergraduate program are equivalent to junior and senior students respectively in the four-year undergraduate program.

<sup>3</sup> For the GEPT elementary, intermediate and high-intermediate level, the perfect score for the listening or reading test is 120; while a combined score of 160 on the listening and reading tests (with neither sub-score lower than 72) is required for passing the first stage of the level. The perfect score and passing score for the speaking or writing test are 100 and 80 respectively (except that the passing score for the elementary writing test is 70).

A total of six administrators (three from each school) were interviewed. Administrators here refer to university presidents, deans of academic affairs and department chairs. A recruitment letter and a consent form were first sent to all the possible candidates before a meeting was arranged for each candidate responding positively.

The students' questionnaire survey was conducted at both schools, followed by interviews with students. A total of 633 students from both schools (305 from School A and 328 from School B) completed the SQ and a total of 32 students (16 from each school) were then interviewed.

The interviews with teachers were conducted after the observation period ended. The eight teachers being observed were automatically invited and six of them were able to be interviewed. Three more teachers from School A and one more teacher from School B were also invited among a few others, totaling a pool of ten teacher interviewees.

#### E. Data Analysis

*Analysis of the SQ data:* For students' perceived impact on their out-of-school practice, perceived impact on their efforts and their worries about the graduation benchmark, a mean score was obtained respectively from the six-point Likert scale.

*Analysis of the interview data:* The interview data were analyzed using the meaning condensation approach (Kvale, 1996). With this approach of data analysis, Kvale (1996, p.194) suggests five steps as in Giorgi's (1975) empirical phenomenological analysis. These five steps were adopted for the present study to analyze the interview data: First, the whole interview was transcribed in full and read through to get a sense of the whole. Second, the researcher determined the natural "meaning units" expressed by the interviewees. Third, the researcher read the interviewees' answers without prejudice and thematized the statements from her viewpoint by stating each natural meaning unit as simply as possible. Fourth, the researcher interrogated the meaning units in terms of the underlying purpose of the present study by asking questions like "What is washback?" and "How was washback generated?" The themes of the meaning units were then addressed with such questions as, "What does this statement tell me about washback?" Fifth, the essential, nonredundant themes of the entire interview were tied together into a descriptive statement to reach interpretation and conclusions.

*Analysis of the observation data:* As mentioned earlier, observation was done to verify what the teachers said in the interviews and to find more themes that might not have emerged in other sources of data. Therefore, all the observation data were analyzed using the meaning condensation approach. The condensed data were then compared to other sources of data, not only to verify the emergent findings but also to provide new insights.

*Analysis of the GEPT and teaching and testing materials:* The two GEPT samples and the teaching materials collected from the two schools were analyzed using the meaning categorization approach (Kvale, 1996). The coding and categorization of the GEPT samples and the collected teaching materials were conducted by the researcher and double checked by a "disinterested peer" of the researcher to ensure their validity.

### IV. RESULTS AND DISCUSSION

To answer the first research question about how administrators, teachers and students perceived the English benchmark policy for graduation, their views on the purposes of the policy will be extracted first, followed by discussions on the related problems with the perceived purposes.

#### A. Perceived Purposes of the English Benchmark Policy for Graduation

Both schools set their English benchmark for graduation higher than the MOE's basic requirement for college graduates in the TVE system. However, if students were unable to pass the test on time, they could take a make-up course as an alternative to fulfill the requirement. What, then, were the real purposes for implementing such a 'flexible' policy? The following views came from interviews with administrators, teachers and students respectively.

*Administrators--The GEPT used for management purposes:* Most administrators mentioned in the interviews one benefit of using the GEPT test; that is, quantification of students' achievements in English, which they thought was important for school management and might even help teachers with their teaching in a sense. In that regard, most administrators did not show much understanding or interest in dealing with students' failure to pass the benchmark before graduation. All that mattered to them was the presentable GEPT passing rates their students could achieve over the years in college.

Obviously, the administrators seemed to be positive and optimistic about the effect of the GEPT test as the graduation benchmark, somewhat consistent with Brown's (2008) observation that administrators' "norm-referenced perspective" tends to keep themselves focused on program-level testing and program evaluation.

*Teachers--The GEPT assumed as an incentive for learning:* Teachers, on the other hand, have "criterion-referenced perspective," which, according to Brown, tends to keep themselves interested in objectives setting, course-level testing, materials development, and delivery of instruction. Supporting evidence could be found in the interview data with some teachers at both schools. For example, Teacher B1 complained that the MOE misled the English education in Taiwan with such a policy, because the MOE had nothing else to check except students' GEPT-taking records. Teacher A5 had the same complaint but still tried to find a balance between teachers' basic concerns and the MOE's requirements, hoping that students could benefit from the policy anyway. Meanwhile, most teachers seemed to agree on the

implementation of a ‘backdoor’ for students, suggesting there was not much confidence perceived by the teachers about students’ success on the GEPT test.

*Students: Taking the GEPT to fulfill social expectations:* Back to Brown’s observation about students’ views on testing, he concludes that all kinds of norm-referenced and criterion-referenced tests will direct the focus, motivations, and efforts of students but only in “the most cynical ways.” Brown’s conclusion seems to suggest that students have no choice but to accept any implementation on them, which seems to be the case with most of the students interviewed at both schools. According to the interview data, most students accepted the implementation of the policy and showed positive attitude towards it. When asked about the intention behind the policy, a list of answers were heard among students, such as “certified English ability is important for future jobs,” “improving English helps globalization,” “preparing for the test helps increasing English proficiency,” “the policy is set for school’s reputation,” etc. Obviously, most of the students agreed on the socially-justified purposes of the English benchmark policy for graduation.

However, negative voices were also heard among a few student participants. For example, some of them saw no need to use English in the future and therefore no need to set a benchmark for them. A few others disapproved the way the GEPT test was used to enhance their English learning and suggested a better environment for learning English was much more important. One student mentioned about the fairness of the graduation benchmark, suggesting there should be different benchmarks for students of different proficiency levels instead of one cutting line for all.

### B. Problems with the Perceived Purposes

In spite of the multiple purposes the graduation benchmark policy were perceived to serve for, the policy at both schools was implemented with a ‘backdoor,’ which seemed to have raised certain doubts about its credibility. Why, after all, was a ‘backdoor’ even necessary here?

*A benchmark set to meet problematic proficiency descriptions:* To understand why a ‘backdoor’ was part of the graduation benchmark policy, it is important to know whether the benchmark was set too high for the students. The SQ survey results show that only a very small portion (2.2%) of students felt the benchmark was easy. Half of the students (53.9%) considered the benchmark was OK and a little less (42.5%) thought it was difficult. Meanwhile, the majority of the teachers and administrators interviewed also considered such a benchmark to be difficult for most of their students. Why, then, did both schools still decide to set the graduation benchmark at this level? Teacher A2 had this typical answer to the question:

#### *Excerpt 1*

The intermediate level of the GEPT is equivalent to the level required for a senior high school graduate<sup>4</sup>. We can not set the benchmark at the elementary level of the GEPT, which is only equivalent to the level required of a junior high school graduate. In that case, the benchmark would be too low!

Apparently, administrators and teachers were misled by the problematic proficiency descriptions at all levels of the GEPT test, as claimed by Shih (2006) on the washback of the GEPT. According to Shih, the LTTC described each level of the GEPT proficiency as equivalent to the fulfillment of each stage of the formal education in Taiwan; for example, the GEPT elementary level was officially described as “the proficiency of those who pass the elementary level is roughly comparable to that of junior high school graduates.” However, evidence shows that some students graduated from senior high schools still could not pass the GEPT elementary test. As a result, Shih suggests that it would probably be more appropriate to describe the corresponding proficiency of the GEPT elementary level as “those who pass the elementary level have mastered the content covered by English textbooks of junior high schools” (p.176).

*A ‘backdoor’ as a necessary evil:* It thus became a thorny problem that many students still could not pass the benchmark before graduation. In order to provide an ‘exit’ for those unable to reach the graduation benchmark, both schools implemented similar make-up measures. What was more, the make-up measure itself even became the major ‘route’ to graduation for most students. For many teachers and administrators, the make-up measure seemed to be a necessary evil and they could not but accept its inclusion in the graduation benchmark policy. Ironically, however, not every student interviewed seemed to appreciate this ‘favor’ for them, though most of the student participants did perceive great difficulties in taking the GEPT intermediate test.

*The ‘pros’ and ‘cons’ of a make-up course as a backdoor:* Some students in School A expressed their dissatisfaction with the make-up course, regarding it as too obvious a ‘backdoor’ to fulfilling the graduation benchmark requirement, which sort of matches the alignment problem Wu (2007) points out in the previous section. Some students in School B even thought that the make-up course was not even necessary. For these very few students, passing the graduation benchmark was certainly important, but what they cared even more was to pass the GEPT test and get the certificate.

However, for the majority of students who perceived great difficulties of the GEPT intermediate test, the make-up measures would certainly be a cure-all. Some even thought of fighting for it if the make-up course should be removed.

*Perceived low stakes of the graduation benchmark policy:* Another underlying problem with the policy was about how teachers and students perceived its stakes, as their perceptions might influence the washback effects of the policy

<sup>4</sup> In brief, the GEPT elementary level used to be described to approximately equal to junior high school graduates’ English proficiency level in Taiwan; the intermediate level to senior high and the high-intermediate level to college. Now, the level descriptions have been modified to a more comprehensive “can do” list as in [https://www.lttc.ntu.edu.tw/E\\_LTTTC/E\\_GEPT.htm#AMenu1](https://www.lttc.ntu.edu.tw/E_LTTTC/E_GEPT.htm#AMenu1).

on teaching and learning (Green, 2007). It was found through interviews that generally low stakes of the graduation benchmark policy were perceived by most teachers and students.

In terms of teachers, most School A teachers perceived the pressure of the graduation benchmark through the university hierarchy, while School B teachers did not, possibly due to the different atmosphere in public and private schools in Taiwan. However, no matter whether the teachers felt pressured by the administration or not, the graduation benchmark did not seem to have brought about any concrete consequences on them.

In terms of students, the stakes of the graduation benchmark were also considerably low as perceived by most students for (1) it was not fatal to fail the exit exam; (2) the backdoor seemed wide open because no schools really wanted to fail their students with the graduation benchmark; (3) in comparison, students put less effort on the English exit exam than on the College Entrance Exam, a relatively high-stakes test in Taiwan.

### C. Washback of the Policy on Teaching

To answer the second research question about the washback effects the policy led to, its influence on teaching will be discussed first followed by discussions on learning.

*Interview data:* The interview data with all the teacher participants at both schools show that teachers did perceive a certain degree of washback of the policy on the following different aspects of teaching: Selection of teaching materials, teaching of test-taking strategies, delivery of the GEPT-related information and teachers' worries about their students' test performance. Overall, the perceived washback on teaching was limited and superficial.

a) *Teachers' selection of their teaching materials.* For the selection of their teaching materials, School A teachers seemed to have perceived more washback than School B teachers. In School A, common syllabi, textbooks, midterm and final exams were arranged by the full-time teachers for the required courses. Part-time teachers must follow the pre-determined course design. In that regard, the full-time teachers thought that they had chosen the teaching materials they considered to be helpful for taking the GEPT test. However, for those part-time teachers, the teaching materials might not have seemed as relevant to the GEPT test as perceived by the full-time teachers.

Unlike School A, School B teachers, no matter full-time or part-time, were free to choose their own teaching materials for the required courses, which were divided into advanced, intermediate and elementary levels. Under such circumstances, very few teacher participants in School B mentioned the influence of the GEPT test on their selection of teaching materials.

b) *Teachers' delivery of test-taking strategies and GEPT-related information.* Some teacher participants also mentioned that in their classes they deliberately spoke of the test-taking strategies that might be useful for taking the GEPT test. Even for Teacher B1, who seemed most unhappy about the policy, the test-taking strategies were also touched upon in his required English classes. Teacher B4, who used the GEPT-like reading and listening comprehension questions in her freshman English classes, explained why she taught test-taking strategies:

#### **Excerpt 2**

When I taught them test-taking strategies, I was trying to deliver a message that, for the GEPT test, some strategies can be used. In fact, quite a few of vocabulary items and grammar points have appeared in the GEPT test repeatedly. For example, past perfect and simple past tense will definitely be tested in the GEPT. So I always remind them of things like that.

In addition to the test-taking strategies, two teachers also mentioned that they would remind students of the GEPT-related information, such as the test date.

c) *Teachers' worries about their students' test performance.* A few teachers felt bound to worry about their students for not being able to pass the graduation benchmark. Teacher B2 had the following representative statement:

#### **Excerpt 3**

OK, I should put it this way. There is no direct influence of the graduation benchmark but, think about it, for a teacher, if there is such a benchmark for the students, will he or she not keep that in mind? Sure, all teachers will keep it in mind.

d) *Limited and superficial washback on teaching.* In response to the obsession with the graduation benchmark, teachers perceived that they have made changes in their textbooks, taught (or mentioned) test-taking strategies or delivered the GEPT test-taking information in their classes. However, when asked if they had made any changes in their teaching methods due to the implementation of the graduation benchmark, no one answered affirmatively. Teacher B3 had the following statement typical of all teacher participants' reaction:

#### **Excerpt 4**

I think my teaching activities have not been influenced [by the GEPT test]. But language learning is nothing more than listening, speaking, reading and writing, right? So your teaching activities should be about practicing these skills. If I say my teaching activities are totally unassociated [with the GEPT test], I think an association is still there. But when I carry out my teaching activities, the GEPT test is never a reason for those activities.

The finding here is consistent with a number of previous washback studies, such as Wall and Alderson (1993) and Cheng (1995), that washback on teaching limits to only the teaching content but teachers' teaching methodology or behavior are rarely changed through a new test.

Interestingly, since the teachers' perceived washback was rather 'superficial' (Cheng 1995), the student participants in both schools perceived nearly no washback on their teachers' teaching in the required English classes. Only very few

student participants remembered that their teachers ever mentioned the GEPT test format or test-taking information in class, reaffirming the trivial washback of the GEPT test on teaching.

*Observation data:* The eight teachers at both schools also showed very little influence from the GEPT test in the classes they were observed, pretty much the same as what they reported in the follow-up interviews. For those who were found to have associated their teaching to the GEPT test, the association was limited to only the announcements of the upcoming internal GEPT mock test. Only Teacher B4 was found to have provided more information than merely a test announcement and she was also found to have used a textbook with reading and listening comprehension questions similar to the GEPT test, though she did not think she had intentionally chosen the textbook for the sake of the GEPT test.

*Analysis of teaching and testing materials:* School A's assigned textbooks were commonly-seen ESL textbooks in the market and were not directly related to the GEPT test, even though School A teachers did perceive an association of their textbooks with the preparation of the GEPT test. Also, a review of their classroom handouts, quizzes and homework assignments shows no direct link with the GEPT test either, with the exception of the sophomores' big test on the GEPT on-line mock tests, which was administered to encourage students to practice the GEPT mock tests on line. Nevertheless, an obvious GEPT influence was found in the listening comprehension test embedded in the mid-term and final exams of the required courses; that is, the listening comprehension test copied exactly the same format of the GEPT test but the contents were based on the assigned materials. The remaining part of the mid-term and final exams, however, did not show direct influence of the GEPT test. The internal test, basically a GEPT mock test administered twice a year, functioned as only a substitute for the external GEPT test as the graduation benchmark.

School B teachers, on the other hand, did not perceive any association of their teaching materials with the GEPT test, though some links were observed: Teacher B3's English learning magazine was undoubtedly tailored for the preparation of the GEPT test but was used mainly for self-study. Teacher B4's reading textbook, as mentioned earlier, did carry some post-reading questions (reading and listening) similar to the GEPT test questions, but the teacher insisted she had chosen the book only out of her own teaching philosophy. The class handouts and quizzes collected from each individual teacher were not found to have a direct link with the GEPT test. However, for some of the common tests in School B, their alignment with the GEPT test could be easily seen without question; for example, the common mid-term and final exams for the freshman required course appeared in the exact format of the GEPT listening test. Also, the two English proficiency tests collected were simply two reshuffled GEPT on-line mock tests, which followed exactly the same format of the GEPT reading and listening tests.

#### D. Washback of the Policy on Learning

This section will continue to look into the washback effects of the policy on learning based on the interview data and student questionnaire survey results.

*Interview data:* Of all the 32 students interviewed at both schools, none of them perceived a high degree of impact of the GEPT test on their learning. More than half of the student participants (19 students) said that they perceived no impact of the GEPT test on their learning, while the rest (13 students) said they perceived only little impact of the GEPT. It is therefore not surprising to find that very few students spent time preparing for the test.

However, when asked about whether they accepted the graduation benchmark policy for English, 18 students showed their acceptance, while only four students were reluctant to accept it. It seems that most of the students were positive about the implementation of the graduation benchmark policy, but why was it so hard for them to be fully engaged? The interview data provide at least three explanations as follows:

a) *Too difficult a goal to accomplish.* As mentioned earlier, the graduation benchmark was too difficult a goal for most of the students to accomplish. For these students, the real problem lay in that they did not know how to prepare for the test but not that they did not want to prepare for the test.

b) *Waiting until the last minute to boost scores.* Some students thought it was still too early for them to start to do anything about it. This is especially true among School A students, since, unlike School B students, they were not required to get a certain score on the first stage of the GEPT intermediate test first before taking the make-up course. Moreover, some students seemed to take cram schools as a last resort to pass the GEPT test, echoing Haladyna et al.'s "test score pollution" (1991, p.2), where score-boosting practices, such as developing a curriculum based on test items, lead to mistaken inferences about ability, which are unethical and should be disallowed. That is probably why some students would rather rely on cram schools than the English courses in school because the regular English classes did not offer enough "score-boosting practices."

c) *Lack of connection between personal needs and the GEPT.* Some students did not feel the connection between taking the GEPT test and their needs, which reduced their motivations for preparing for the test. For example, Student A7 felt this way:

##### **Excerpt 5**

I think the most important thing for learning English is to learn to speak, instead of taking written tests. But in Taiwan, our tests focus only on written tests, like the GEPT and our mid-term and final. The College Entrance Exam is a written test too, focusing on grammar. I think even native speakers may not be able to get good grades on such a test, because it is too hard. I think these kinds of tests (written tests) are of little help to our English learning. Even if you get high scores on the tests, it doesn't mean that you can speak English well.

Student A7's emphasis on spoken English, like many other students' views, seems to agree with Yang's (1992) BALLI study on Taiwanese students that most students believed speaking English well would bring in many opportunities. In a similar vein, some highly-motivated students even said that they study English purely out of their own interest, not because of the GEPT test. In other words, the GEPT test had no impact on them, because they would study English on their own no matter whether there was a graduation benchmark or not.

*SQ survey results:* A total number of 633 students participated in the SQ survey. As 27 student participants indicated in the SQ survey that they already passed the benchmark before entering college, these students were excluded from the subsequent statistical analysis of students' perceived impact of the English exit exam on the three dimensions (*students' perceived impact on their out-of-school practice, students' perceived impact on their efforts* and *students' worries*). The descriptive statistics of the remaining 606 students are presented in Table 1, showing that the mean score for *students' worries about the graduation benchmark* (3.66) is higher than the other two dimensions (3.37 and 1.33) on a six-point scale.

TABLE I  
DESCRIPTIVE STATISTICS FOR STUDENTS' PERCEIVED IMPACT OF THE ENGLISH EXIT EXAM ON THE THREE DIMENSIONS  
(EXCLUDING STUDENTS WHO ALREADY PASSED THE BENCHMARK BEFORE COLLEGE)

	N	Minimum	Maximum	Mean	SD
Out-of school practice	606	1.00	4.00	1.33	.40
Efforts	606	1.00	6.00	3.37	.93
Worries	606	1.00	6.00	3.66	1.06

## V. LIMITATIONS OF THE PRESENT STUDY

Like most quantitative and qualitative studies, this present study does have its limitations. First, due to the elusive and complex features of the washback phenomenon (Spratt, 2005), it is often problematic for a questionnaire survey to capture the washback effects caused by the test in question, such as what happened in the SQ survey, where high-achieving students' responses to their perceived washback of the English benchmark for graduation were not clearly separated from their eagerness to learn more English, as revealed by the qualitative data. As a result, there are certain difficulties gathering washback-related data with one single approach. Triangulation of different sources of data is essentially important to a washback study given that the elusiveness and complexity of washback always exist in each and every situation.

Second, owing to the willingness of the teachers being observed, some classes were observed only once or twice for the present study. It is inevitable that these short-term observations might have led to certain bias in the results and analysis. However, this shortcoming has been noticed and the observation data have been triangulated with the related interview data and teaching and testing materials to ensure their validity.

Third, one common problem with the qualitative analysis is their generalizability. However, this problem has been taken care of by so called "thick description" (Geertz, 1973) in order to draw the research findings from an "extensive and careful description of the time, the place, the context, the culture" (p.241), which could be used for other researchers' references if the research contexts are similar.

## VI. CONCLUSION

This investigation has confirmed a social consensus about the implementation of the graduation benchmark for English in Taiwan's higher education, but it has also found that the policy didn't quite achieve the MOE's goals to increase college students' overall English proficiency and their global competitiveness by looking at its washback effects. As the washback on teaching was limited and superficial, including only the use of the GEPT mock tests, adoption of the GEPT test format for teacher-made tests, teaching of test-taking strategies, delivery of the GEPT-related information and teachers' worries about their students, and the washback on learning was also very little and mostly negative, such as test-induced fear, pressure, anxiety and frustration, it is quite clear that the policy did not bring about any significant and meaningful changes in the educational settings as intended.

However, according to the follow-up plan outlined in *Challenge 2008: National Development Plan Summary Report of Main Achievements*, the MOE still decided to keep encouraging universities and colleges to set their graduation thresholds using standardized English tests. Students' passing rates would also continue to be one of the indicators for college evaluation, in an attempt to "globalize the tertiary education in Taiwan" as before.

As reflected in the present study, the MOE of Taiwan might have needed to listen to different stakeholders' voices before such an important decision was made, especially those from teachers and students, who were directly responsible for the intended results. After all, this test-oriented English benchmark policy for graduation has proven not as effective as what the MOE originally expected. To achieve the MOE's goal for internationalizing Taiwan's higher education and improving college students' English proficiency, the graduation benchmark might have to relate more purposefully to the subsequent policies, and tests, especially standardized proficiency tests, might no longer be a panacea as traditionally believed in the Asian context.

APPENDIX A. TEACHERS INTERVIEWED/OBSERVED

School A					
Teachers	Gender	Full-time/Part-time	Teaching experience	Interviewed	Observed
A1	F	P	4 yrs.	✓	
A2	F	F	8 yrs.	✓	✓
A3	F	F	14 yrs.	✓	✓
A4	F	F	24 yrs.	✓	
A5	F	F	18 yrs.	✓	
A6	M	P	4 yrs.		✓
School B					
Teachers	Gender	Full-time/Part-time	Teaching experience	Interviewed	Observed
B1	M	F	24 yrs.	✓	✓
B2	F	F	24 yrs.	✓	✓
B3	F	F	24 yrs.	✓	✓
B4	F	F	5 yrs.	✓	✓
B5	F	P	25 yrs.	✓	
B6	F	F	20 yrs.		✓

APPENDIX B. ADMINISTRATORS INTERVIEWED

School A			School B		
Administrators	Gender	Position	Administrators	Gender	Position
A1	M	President	B1	M	Dean of Academic Affairs
A2	M	Dean of Academic Affairs	B2	M	Ex-Department Head of Applied English
A3	F	Ex-English Language Section Coordinator	B3	F	Department Head of Applied English

APPENDIX C. STUDENTS INTERVIEWED

School A				School B			
Students	Gender	Majors	Years of study	Students	Gender	Majors	Years of study
A1	M	Electronic Engineering	1	B1	F	Electronic Engineering	1
A2	M	Environmental Engineering	1	B2	F	Industrial Design	1
A3	F	Industrial Design	1	B3	F	Chemical Engineering	2
A4	M	Mechanical Engineering	1	B4	F	Chemical Engineering	2
A5	F	Chemical Engineering	1	B5	M	Organic and Polymeric Materials	2
A6	M	Vehicle Engineering	2	B6	M	Materials Engineering	2
A7	M	Electronic Engineering	2	B7	M	Mechanical Engineering	2
A8	M	Vehicle Engineering	2	B8	M	Architecture	3
A9	M	Mechanical Engineering	2	B9	F	Mechanical Engineering	3
A10	M	Business Management	2	B10	M	Chemical Engineering	3
A11	M	Electrical Engineering	4	B11	M	Electrical Engineering	3
A12	M	Electrical Engineering	4	B12	M	Civil Engineering	3
A13	M	Electronic Engineering	4	B13	F	Industrial Management	3 (2-yr.)
A14	M	Electrical Engineering	4	B14	M	Mechanical Engineering	3 (2-yr.)
A15	M	Chemical Engineering	4	B15	F	Industrial Management	4
A16	M	Electronic Engineering	4	B16	M	Materials Engineering	4

## REFERENCES

- [1] Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- [2] Brown, J.D. (2008). Effects matrix for administrators', teachers', and students' views on testing for ELI at UHM. Observation notes presented at the annual meeting of the 30<sup>th</sup> Language Testing Research Colloquium (LTRC), June 23-28, Hangzhou, China, 2008.
- [3] CEPD (Council for Economic Planning and Development). (2008). Challenge 2008: National development plan summary report of main achievements. Taipei: Council for Economic Planning and Development, Executive Yuan.
- [4] Chen, J. F., Warden, C. A., & Chang, H. (2005). Motivators that do not motivate: The case of Chinese EFL learners and the influence of culture on motivation. *TESOL Quarterly*, 39(4), 609-633.
- [5] Cheng, L. (1995). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11, 38-54.
- [6] Geertz, C. (1973). Interpretation of cultures. New York: Basic Books.
- [7] Giorgi, A. (1975). An application of phenomenological method in psychology. In A. Giorgi, C. Fischer & E. Murray. (Eds.), *Duquesne studies in phenomenological psychology II*. Pittsburgh, PA: Duquesne University Press.
- [8] Green, A. (2007). IELTS washback in context: Preparation for academic writing in higher education. *Studies in language testing* (Vol. 25). Cambridge: Cambridge University Press.
- [9] Haladyna, T.M., Nolen, S.B., & Haas, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 20-25.
- [10] Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.
- [11] Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
- [12] Shih, C. (2006). Perceptions of the General English Proficiency Test and its washback: A case study at two Taiwan technological institutes. Unpublished doctoral dissertation. University of Toronto, Toronto.
- [13] Shohamy, E. (1993). The power of test: The impact of language tests on teaching and learning. *NFLC Occasional Papers*. ED 362-040
- [14] Simpson, A. A. (2007). Language and national identity in Asia—Taiwan. In A. A. Simpson (Eds.), *Language and national identity in Asia* (pp. 235-259). Oxford: Oxford University Press.
- [15] Spolsky, B. (2004). *Language policy*. Cambridge: Cambridge University Press.
- [16] 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.
- [17] Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.
- [18] 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*, 574-586. Taipei: Crane Publishing Company.
- [19] Wu, J. R. W., & Wu, R. Y. F. (2007). Using the CEFR in Taiwan: The perspective of a local examination board. Paper presented at the fourth annual EALTA conference, Sitges, Spain, 2007.
- [20] Yang, N. D. (1992). Second language learners' beliefs about language learning and their use of learning strategies: A study of college students of English in Taiwan. Unpublished doctoral dissertation. University of Texas, Austin.

**Hsiu-yu Chu** is currently an associate professor of the General Education Center at Ming Chi University of Technology in Taiwan. She received a Master's degree in linguistics from University of North Carolina at Chapel Hill and her PhD in TESOL from National Taiwan Normal University. Her research interests include washback studies, language testing and second language acquisition.

**Hsi-nan Yeh** is currently an associate professor of the Department of English at National Taiwan Normal University. He received a Master's degree in English from National Taiwan Normal University and his PhD in curriculum and instruction from University of Texas at Austin, USA. His research interests include EFL program planning and evaluation, language testing and teacher development.