

Traditional and Non-Traditional Test Preparation Practices: Learner Performance and Perspectives

Yi-Ching Pan

(yichingpan@yahoo.com.tw)

National Pingtung University, Taiwan

Abstract

The integration of beneficial washback of tests within instruction is a major feature in arguments around test validity. Beneficial washback occurs when a test encourages instruction that leads to improvement in target abilities, not just instruction focused on passing the test. Issues arise in this regard in the context of English graduation benchmark remediation classes in Taiwan. In order to enhance students' English proficiency, around 90% of Taiwanese tertiary education institutions use the General English Proficiency Test as a graduation requirement. To ensure that students achieve the graduation benchmark, 80% of technical universities and colleges offer remedial instruction to at-risk students. Against this background, the study reported here explored the impact of test preparation courses on learners' test performance. It also investigated learners' preferences and perceptions regarding the effectiveness of two approaches to test preparation, namely a traditional and a non-traditional approach. Pre- and post-test scores as well as questionnaire and interview data were gathered from 72 Taiwanese university students. Half of these attended non-traditional test preparation classes that included communicatively-oriented activities, whereas the other half attended traditional classes with test-oriented activities. The non-traditional group achieved significantly greater score gains in listening, while the traditional group did better in reading. Furthermore, test-oriented instruction was reported by the subjects to be more efficient than communicatively-oriented instruction in terms of enhancing their English proficiency.

1 Introduction

Tests are believed to have effects on both the learners preparing for them and the teachers assisting such learners with their preparation (Spolsky, 1999). The influences of testing on teaching and learning are commonly referred to as “washback effects” in the field of language testing (Bailey, 1996; Hughes, 2003). In recent years, analyzing the consequences of test use has become an important aspect of test validation, and washback is one dimension of such testing consequences in the classroom. Ensuring beneficial washback of tests onto instruction is therefore a major aspect of test validity. Such beneficial washback occurs when a test encourages instruction that leads to improvement in target abilities, rather than just instruction focused on passing the test (so-called “test preparation”).

English graduation benchmark remediation classes in Taiwan currently reflect the above scenario. In order to enhance students' English proficiency, around 90% of Taiwanese tertiary education institutions have adopted the General English Proficiency Test (GEPT) as a graduation requirement. The percentage of students passing with an equivalent of basic English proficiency increased from 32% in 2006 to 56% in 2009 (Taiwan Ministry of Education, 2012), but further statistics from the same source reveal that a substantial percentage of students remain unable to

meet the graduation benchmark. To ensure that such students will achieve the graduation benchmark, allowing them to graduate, 80% of technical universities and colleges offer remediation instruction to at-risk students.

According to its developer, the GEPT is designed to promote an emphasis on communication skills in formal English instruction (Wu, 2012). Preparing students for this standardized test is therefore often considered ethically problematic due to concerns over issues of narrowing the curriculum, overemphasizing test-strategy instruction, and mechanical practice. As learners are subject to the most direct impact of testing (Wall, 2000), increasing attention has been devoted to the direct influence of test preparation on learners. Many such studies have investigated learners' out-of-class test preparation practices, test-taking strategies, and test scores (e.g. Damankesh & Babaii, 2015; Stoneman, 2006; Zhan & Andrews, 2014). What remains unclear is the extent to which learners regard in-class test preparation practices as effective in enhancing their English proficiency. To address this gap in the research, the study reported here aimed to explore the impact of test preparation practices on learners' test performance. In addition, learners' preferences and perceptions regarding the effectiveness of two approaches to preparation for the GEPT were investigated, namely a traditional and a non-traditional test preparation course. Three research questions were addressed in this study:

- (1) What is the impact of traditional and non-traditional test preparation courses, respectively, on learners' test performance?
- (2) What factors, if any, are perceived by traditional and non-traditional learners, respectively, as leading to failure on the practice GEPT test?
- (3) What are the perceptions of traditional and non-traditional learners, respectively, regarding the effectiveness of test preparation activities in enhancing their English proficiency?

2 Literature review

2.1 Washback and validity

Researchers offer different perspectives of the role of washback in validity theory. One view (Davis, 1997; Mehrens, 1998; Popham, 1997) is that consideration of the ways in which test use affects validity overburdens the concept. According to this argument, a unified theory of construct validity is a direct result of the accuracy of score-based inferences, implying no need to investigate further when validating a test. The opposing argument (e.g. Cheng, 2014; Kane, 2006; McNamara & Roever, 2006) is that test validation is not utilized to examine the validity of the test itself, or of test scores per se, but rather to determine whether or not interpretations, inferences, and decisions based on test use are justified.

The present study is sympathetic to the latter view that the consequences of test use should be considered as an aspect of validity research. Furthermore, the effect of a test on a classroom is a type of test consequence; thus, consideration of washback is related to test validity. Validity is usually defined as the degree to which a test measures what it claims to measure, and any mismatch between a test construct and what is taught, learned, or even perceived may be a threat to test validity. Taking the GEPT as an example, if test preparation practices improve test scores by enhancing communicative competence, which is what the GEPT claims to measure, such practices pose no threat to construct validity. On the other hand, if unethical practices, such as the enhancement of test-taking skills, often adopted in traditional test preparation courses, are used to help learners improve their test scores, this may harm test validity.

Given the large number of GEPT test takers (4.3 million) since its launch in 2002 (Wu, 2012), it would be useful to know whether or not in-class GEPT preparation activities have any effect and, if so, which are most effective from the viewpoint of learners, who are the major stakeholders. Furthermore, the issue of whether or not test preparation practices that comprise both test-oriented and communicative teaching are considered a threat to test validity remains unanswered in the field of language testing.

2.2 *Research on test preparation practices*

During the past two decades, an increasing number of washback studies have been conducted within the field of teaching English as a foreign language, but studies focused on actual learners remain limited (Cheng, 2014; Wall, 2000). Among the few learner-centered washback studies, test preparation is a general issue that has been investigated for three types of tests, namely (1) traditional, multiple-choice, large-scale high-stakes tests (Damankesh & Babaii, 2015); (2) local or international English proficiency tests such as The General English Proficiency Test (GEPT), the Test of English for International Communication (TOEIC) and the International English Language Testing System (IELTS) (Gan, 2009; Green, 2007; Pan, 2014; Stoneman, 2006); and (3) new or modified tests (Wall, 2005; Xie, 2013). The following sub sections present a review of learners' test preparation activities and the effects of test preparation practices on learning outcomes, respectively. This overview is given to facilitate an understanding of test effects on learners, and to highlight the gap in the research that requires further attention.

2.2.1 *Learners' test preparation activities*

Learning activities utilized by learners as test preparation practices are similar across the three types of tests mentioned above. Thus, the test type does not influence learners' strategies for learning English or preparing for the test. Stoneman (2006) investigated how 655 learners in Hong Kong prepared for a local university graduation exam and the IELTS test. Traditional learning strategies and exam preparation methods, such as reading textbooks, memorizing vocabulary and idioms, going through previous exams, and relying on test preparation books, were prevalent among the research subjects in preparing for the different tests. Pan (2014), in turn, explored the learning activities of 457 Taiwanese university students in different years of study and of different levels of proficiency in preparing for the exit test (either the GEPT or the TOEIC). The findings showed that higher proficiency learners adopted a higher frequency of language skill-building activities to prepare for the test, although all subjects favored test-related activities. Their test-related activities included online practice tests and mock tests assigned by their instructors. Their language skill-building activities included reading magazines, newspapers, and novels, listening to English broadcast programs, keeping an English diary, talking with foreigners, joining English conversation clubs, and practicing English with teachers. Damankesh and Babaii (2015) investigated how 80 Iranian high school students prepared for their high-stakes final exam. The primary test preparation strategies reported were studying only what was tested, and memorizing and reviewing past papers. As teachers can "determine to a greater or lesser extent whether to allow washback to operate, what areas it should operate in, and how" (Spratt, 2005, p. 24) and as learners' in-class experience certainly interacts with their test preparation methods (Zhan & Andrews, 2014), it is worthwhile to investigate whether learners' perceptions of the effectiveness of teaching activities in test preparation classes are similar to those related to their out-of-class test preparation practices.

2.2.2 *The effects of test preparation practices on test performance*

Wall (2000) contends that "[w]hat is missing... are analyses of test results which indicate whether students have learnt more or learned better because they have studied for a particular test" (p. 502). Wall and Horak (2006) also claim that much of the research regarding washback or impact has focused less on investigating test effects on products than on teaching practices, that is, the outcome or quality of learning. The findings of a handful of studies on the effects of test preparation practices remain inconclusive. Whereas some suggest that test preparation practices can promote language learning, others suggest little or no significant impact. Elder and O'Loughlin (2003), for example, examined the relationship between intensive test-driven English language study and IELTS score gains after 10 to 12 weeks of instruction among 130 students in New Zealand. They found significant gains in listening, but not reading. A range of factors appeared linked

to test score improvements, such as personality, motivation, confidence, and exposure. Xie (2013), in turn, investigated 1,003 Chinese university students, and discovered that test takers who did eight weeks of preparatory practices by narrowing the preparation content, especially drilling, obtained higher scores on College English Test Band 4 than did those who prepared less or not at all. Conversely, Green (2007) investigated whether test preparation classes significantly impacted IELTS writing scores, showing that such courses offered “no clear advantage” (p. 75) in terms of IELTS grammar and vocabulary test performance across different groups over a four to 14-week period. Green also found that score gains occurred primarily among two groups of learners, namely those who planned to take the test again and those who had low initial writing scores. Similarly, Gan (2009) investigated 141 Hong Kong university students, and found no significant differences in exit IELTS scores between students who had taken a 20-hour preparation course and those who had not. However, those who attended a preparation course showed greater improvement from their original low scores. The aforementioned studies suggest that learners’ proficiency, learners’ style, time of exposure, and motivation are possible factors affecting improvement. While such studies have added to our understanding of score gains on various English proficiency tests, two issues have not yet been seriously examined, and are therefore the focus of this study: (1) the issue of test improvement over a longer period of time, such as over a semester rather than a short intensive test preparation course, as was the case in most studies, and (2) the issue of the relationship between test-related factors (test difficulty, test stakes, test content) and test-unrelated factors (English proficiency, motivation, time of exposure, teaching activities) in explaining student failure on a test.

3 The General English Proficiency Test

In 1999, Taiwan’s Ministry of Education commissioned the Language Training and Testing Center (LTTC) to develop the GEPT with several objectives in mind, one of which was to encourage the study of English in a communicative manner. The GEPT was developed at least partially in response to a belief among educators that Taiwanese learners of English were generally unable to communicate in English due to an outdated approach to teaching that placed too much emphasis on grammatical accuracy. As of 2014, more than 4.3 million learners have taken the GEPT (Wu, 2012).

The GEPT covers five levels of proficiency, namely elementary, intermediate, high intermediate, advanced, and superior. It is designed to test all four language skills, namely listening, speaking, reading, and writing. Test takers must pass two stages of the test in order to receive a GEPT certificate. The first stage consists of a paper-and-pencil test targeting listening and reading skills, with items in a multiple-choice format. The second comprises a recorded session targeting speaking and writing skills. The elementary GEPT is set as an exit requirement by the majority of Taiwanese universities of technology. According to the LTTC (2015), test takers who pass the elementary level have a basic ability in English and can understand and use rudimentary language required in daily life.

4 Methodology

4.1 Subjects

Subjects for this study were recruited at a university of technology that requires students to fulfil an English graduation requirement before they graduate. Students may choose from an array of English proficiency tests, namely the GEPT, the TOEIC, the Test of English as a Foreign Language (TOEFL), or the IELTS, and have to achieve a score equivalent to A2 of the Common European Framework of Reference for Languages (CEFR). Students who do not meet the graduation requirement in their fourth year are required to take 54 hours of remedial English classes over an 18-week period and to meet the requirements set for the course in order to graduate. They are allowed to drop out during the course if they achieve a passing standard of the required test.

For the purposes of this study, 77 fourth-year students registered for such remedial English classes in the Fall semester of 2014 were randomly assigned to two remedial classes by the university, with 40 students in the traditional remediation (TR) class and 37 in the non-traditional remediation (NR) class. Of the 77 subjects, one dropped out of NR class, and four out of the other, having passed the elementary GEPT, leaving 36 subjects in each class. Table 1 presents their gender information.

Table 1. Gender of subjects

	NR class N=36	TR class N=36
Male	11	28
Female	25	8

4.2 The traditional and non-traditional remedial classes

In order to compare the effectiveness of a traditional and a non-traditional approach, two types of GEPT-based test preparation classes were designed and conducted by the same instructor, who was the researcher in this study. The NR class aimed to help students to prepare for the elementary GEPT and simultaneously to improve their communicative skills, while the TR class aimed solely to help students prepare for the test. Appendix 1 lists the major activities conducted in the two classes. The communicative activities are marked with “✓”, while the test-oriented activities are marked with “◇”.

The NR class included both test-oriented and communicatively-oriented activities. In addition to lectures and the practicing of test questions on vocabulary, grammar, listening, and reading, communicatively-oriented activities took the form of teacher-student interactions, and pair and group discussion. For example, before practicing listening test questions, the instructor would ask the NR subjects what they saw in a given picture, or require them to practice in pairs by developing a conversation based on a given picture with the aid of a vocabulary handout. Subjects were also asked to discuss, in pairs or groups, pre-reading, reading, and post-reading questions and longer reading comprehension questions. With much of the time allotted to communicatively-oriented activities, test-oriented activities in the NR class focused solely on the explanation of confusing and difficult test questions, which were explained in Chinese.

Test-oriented activities in the TR class included traditional teacher-dominated lectures and translation of reading materials and explanation of test questions in Chinese. Individual practice and pair and group discussion were also conducted in the TR class, during which subjects translated reading materials or test questions from English into Chinese, or found answers to test questions in the medium of Chinese. These activities were focused on test preparation, and could not be characterized as communicatively-oriented, neither did they create knowledge or promote critical thinking.

One activity common to both classes was a weekly culture talk of three to five minutes as a warm-up activity at the start of class. For the purposes of this article, the activity of sharing foreign cultural and life experiences was categorized as a communicatively-oriented activity. The main goal of the textbook, “Off We Go 1”, is to provide students with a fundamental grasp of English vocabulary, grammar, and sentence structure, drawn from reading materials on a variety of topics. It also provides supplementary GEPT listening and reading practice tests for those who wish to take the test.

4.3 Instruments

The same practice elementary GEPT test, published by the LTTC, was utilized as the pre- and post-test for this study. The data for the present study were those reflecting receptive skills only (i.e. listening and reading), as the subjects were not required to submit their scores for productive

skills (i.e. speaking and writing). As test papers and answers were not given to the students until they had finished the post-test, the ceiling effect was avoided. In addition, by using the same test paper for pre- and post-testing, the study intended to compare the two groups in terms of score gains. The 30 items in the listening subset and 35 items in the reading subtest were multiple-choice questions. According to the LTTC, the passing standard is a total score equal to or above 160, with each subtest score (listening and reading) no lower than 72 out of 120.

In addition to the elementary GEPT listening and reading subtests, a questionnaire was used to investigate subjects' perceptions of (1) the effectiveness of the various teaching activities in the remedial classes, and (2) the test-related and test-unrelated factors contributing to their failure on the elementary GEPT post-test. The questionnaire comprised 16 items targeting (1) above and 14 items targeting (2) above. Subjects were required to indicate their level of agreement with a given statement on a 5-point Likert scale, and were welcome to express their opinions fully at the end of the questionnaire. The questionnaire was presented in Chinese to facilitate its completion and to put the subjects at ease (Johnson & Christensen, 2000). In terms of internal consistency and validity, the Cronbach's α r for the 16 items on the effectiveness of teaching activities was 0.86, and that for the 14 items on factors affecting failure was 0.73 (Tavakol & Dennick, 2011, suggest that acceptable values range from 0.70 to 0.95).

Finally, three students from each of the NR and TR classes volunteered to be interviewed individually for 10 to 15 minutes each, giving them an opportunity to express their own viewpoints. The interview data were audio-recorded and transcribed to identify the key points, and were used to better understand the data from the questionnaire.

4.4 Data collection procedures

The elementary GEPT pre-test took place in the first week of the remedial classes, and the post-test 17 weeks later during the last week of classes. Each test took approximately 50 minutes to complete. The questionnaire forms were distributed during the last week of classes. The 72 subjects who completed the questionnaire and took both the pre- and post-tests were rewarded with a token of the researcher's appreciation. Three subjects from each of the two classes volunteered to participate in the individual interviews.

4.5 Data analysis

PASW Statistics Version 18 was used to analyze the quantitative data and test scores. Descriptive statistics were utilized to calculate frequency distributions, means, and standard deviations for the GEPT questionnaire scores. Inferential statistics (independent t-tests) were used to check for levels of statistical significance ($p < 0.05$). The interview data were interpreted in terms of the quantitative data.

5 Results

5.1 Impact of traditional and non-traditional test preparation on test performance

A paired-samples t-test was used to compare the mean pre- and post-test scores for the TR and NR subjects. Both groups performed statistically significantly better on the post-test than the pre-test in both listening and reading (see Table 2). Thus, both types of remedial classes appeared to be effective. However, the passing percentage for the TR and NR classes (13 NR students, 36%, and 5 TR students, 1%) was not satisfactory.

As Table 2 also shows that the subjects' initial levels of English proficiency were dissimilar, with NR subjects' total scores statistically significantly higher than those of TR subjects ($t(70)=3.925, p<.05$), this does not show which type of remedial instruction promoted better test performance. Therefore, an independent t-test was conducted to compare as the two groups in terms of score gains. The NR subjects improved significantly more than their TR counterparts in

listening ($t(70)=3.085, p<.05$), while the TR class improved significantly more than the NR class in reading ($t(70)=2.136, p<.05$). Given the large effect size for listening ($d=.737$) and the medium effect size for reading ($d=.5$), it appears that the higher frequency of communicatively-oriented activities may have promoted the NR subjects' listening comprehension, while the higher frequency of test preparation activities may have enhanced the TR subjects' reading abilities. With regard to this finding, it is not surprising that 27% of the NR subjects ($N=10$) suggested on the questionnaire that more class time ought to be allotted to lectures on grammatical issues to help them analyze and understand lengthy sentences on the test.

Table 2. GEPT means scores and standard deviations at pre- and post-testing

	TR (N=36)						NR (N=36)					
	Listening		Reading		Total		Listening		Reading		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Pre-test	54.56	12.25	47.56	13.63	102.11	23.54	57.33	15.68	64.56	13.50	121.89	23.23
Post-test	63.19	12.86	61.25	18.64	124.44	27.29	76.83	13.88	71.39	16.82	148.22	24.05
	t	df	t	df	t	df	t	df	t	df	t	df
Paired t-test	7.09	35	3.38	35	7.46	35	3.92	35	5.47	35	6.22	35
Sig	.000		.002		.000		.00		.00		.00	

5.2 Factors perceived to account for test failure

The percentage of both TR and NR subjects that passed the GEPT post-test was unsatisfactory (13 NR students, 36%; four TR students, 1%). Specifically, the average total post-test scores (TR=124.44, NR= 148.22) did not meet the threshold of 160 set for the graduation requirements. When asked what may have led to their failure on the GEPT post-test, 55 participants out of these 72 replied on the questionnaire. 23 NR subjects and 32 TR subjects reported both test-related and test-unrelated variables (see Table 3). The test-unrelated variables included limited proficiency in vocabulary, phrase structure, and grammar, which were reported as major concerns in their failure, reflecting the three highest means ($M=4.37\text{--}4.10$). The following variables were related to test difficulty, including listening speed, lengthy listening/reading context, and topic unfamiliarity, with means ranging from 4.06 to 3.09. The factors with the three lowest means ($M=3.00\text{--}2.69$) included both test-related and test-unrelated variables, namely insufficient availability of test preparation courses, low interest in teaching activities, and little benefit associated with GEPT certification. These findings indicate that English proficiency and test difficulty were the two major factors leading to subjects' GEPT post-test failure.

The interview data reflected similar reasons for subjects' failure. For example, TR1 reported that she usually worked very hard at the beginning of the test, but this did not last long due to boredom and frustration: "I felt so frustrated and bored when seeing so many words that I did not know, and I gave up." Indeed, TR2 reported that he fell asleep during the test: "There are just too many new and difficult words. They probably know me, but I don't know them, unfortunately." NR1 frowned at her low score, saying, "I don't know how to memorize the words and phrases, even though I tried."

An independent t-test was conducted to compare the TR and NR subjects' perceptions regarding factors related to their GEPT post-test failure. The only statistically significant difference was for a test-unrelated factor: a significantly larger number of TR subjects (44%, as opposed to 17% of NR subjects) thought they failed because what was taught in class did not motivate their English learning ($t(70)=3.005, p<.05$). The effect size for this variable was $d=.72$. This finding suggests that test-oriented learning, focusing on practicing test item types, rather than on developing language proficiency, may have been what the TR subjects needed to attain the required test scores, but they appeared to tire of such learning, and their motivation was not enhanced in this narrower test-based remedial course.

When asked about further suggestions for remedial classes, 36% of NR students (N=13) reported liking the way they were taught in class, and that no changes were required. They offered positive comments on the questionnaire, such as “vocabulary size has become larger,” “no longer so afraid to speak in class,” “more time spent with English,” “grammar is better now,” and “learned how to answer test questions.” Several TR students, however, wished the remedial classes to be more interesting, as reflected by the following questionnaire comments: “life English,” “fewer tests,” “more listening,” “too much grammar,” and “expected a foreign teacher.” In the interviews, TR1 said, “What I have learned is tests only. It’s a bit meaningless that I am not able to speak English with foreigners.” While, TR2 and TR3 expressed their appreciation for the teacher’s hard work, they also complained of being bored with the lengthy explanations of grammatical features. TR3 suggested that “it would be nice to learn ‘a theme of vocabulary’ every week. English self-introduction, for example. Or teaching what is often used and needed in daily life.” These findings suggest that traditional teacher-centered teaching activities focusing on test preparation and grammar translation did not appeal to the lower-achieving subjects in the TR class.

Table 3. Factors perceived to relate to GEPT post-test failure

Test-related variables	TR subjects (N=32)		NR subjects (N=23)	
	M	SD	M	SD
Test difficulty				
Fast listening speed	4.06	.87	3.78	.95
Lengthy listening/reading context	3.75	.91	3.74	1.17
Topic unfamiliarity	3.59	.94	3.57	1.08
Test content				
Low learning motivation with GEPT content	3.50	.95	3.22	.95
Test stakes				
Little benefit associated with GEPT certification for further studies or job hunting	2.75	1.19	3.09	1.16
Test-unrelated variables				
Proficiency				
* Limited size of vocabulary	4.22	.65	4.52	.51
* Limited size of phrases	4.16	.72	4.43	.50
* Limited knowledge of grammar	3.91	.96	4.17	.83
Limited pronunciation skills	3.59	.97	3.74	1.01
Learning style				
Test phobia	3.28	1.27	3.13	1.10
Low interest in English	3.66	1.20	3.57	1.08
Limited time devoted to English study	3.78	.75	3.70	1.06
Resources				
Not enough test preparation courses not offered	2.88	1.12	2.43	.84
* Low motivation in learning due to what is taught in class	3.44	.80	2.61	1.03

5.3 Perceptions of the effectiveness of test preparation activities in enhancing English proficiency

In general, traditional teacher-centered test preparation activities, including lectures and the practicing of grammar, vocabulary, listening, and reading test questions, were reported by both TR and NR subjects to be the most effective methods in enhancing their English abilities, with means ranging from 4.26 to 4.07. Communicatively-oriented activities, in contrast, such as pair and group discussion, watching English films and completing short film reflections, and the sharing of for-

eign cultural experiences, were considered to be the least effective, with means ranging from 3.76 to 3.43. When asked why they favored test-oriented activities over communicatively-oriented activities, many interviewees attributed this to their low proficiency in vocabulary and grammar. For example, TR3 reported, "I could talk a lot if I were allowed to speak in Chinese. My poor vocabulary prevented me from expressing what I meant to say." Similarly, NR2 said, "It was kind of wasting time doing pair practice because my partner did not want to communicate in English, and we ended up chatting in Chinese ... It's too troublesome to look up for words and phrases with my mobile phone when talking with someone in English." These data suggest that the lower-proficiency subjects did not benefit from opportunities for the communicative exchange of information and interaction that were presented in the remedial test preparation classes.

An independent t-test (see Table 4) was conducted to determine the significance of any differences in TR and NR subjects' perceptions with regard to the effectiveness of teaching activities in the two types of remedial classes. Two major findings were revealed. First, the NR students had significantly more positive perceptions than their TR counterparts in terms of a number of test-oriented activities perceived as improving their English abilities: (1) lectures and grammar practice ($t(70)=2.791, p<.05$); (2) lectures and practice of listening test questions ($t(70)=2.316, p<.05$; $t(70)=2.531, p<.05$); (3) lectures and practice of reading test questions ($t(70)=2.295, p<.05$; $t(70)=3.289, p<.05$). The NR subjects expressed mixed opinions regarding grammar teaching on the questionnaires. Ten NR subjects wished for more grammar lectures, while nine thought that their grammar knowledge had increased due to the remedial classes. As for the NR subjects' preference for test-related practice, certain explanations emerged from the interviews. NR2 reported that he was a passive learner, and would not have spent time on English study if he had not been required to do test-practice in class: "In addition, doing tests with my hand moving would help me not to fall asleep so easily. Honestly speaking, I don't like tests very much, but this way of teaching suits my personality." Furthermore, NR3 stated that the major purpose for her to attend class was to do a lot of practice tests: "This is a test preparation class, not a regular English class."

Second, the TR subjects had significantly more positive perceptions than their NR counterparts regarding the effectiveness of sharing foreign culture and life experiences in improving English proficiency ($t(70)=2.036, p<.05$). This finding suggests that the TR subjects were bored with traditional test preparation activities and were particularly interested in the occasional culture-related discussions to motivate their English learning. TR1 provided the following reason: "The positive influence of this remedial class is that the teacher shares her study-abroad experience." When asked whether the culture share activity motivated her English learning, TR1 said "Yes, we've stayed all our life in Taiwan, but we've never had the experience to see what's like outside the world. Thanks to culture share, I can gain some more knowledge."

Table 4. Perceptions of the effectiveness of teaching activities in improving English proficiency

	TR subjects (N=36)		NR subjects (N=36)	
	M	SD	M	SD
Test-oriented activities				
Lectures of vocabulary instruction	4.17	.60	4.36	.63
*Lectures of grammar and practice of grammar exercises	4.03	.77	4.47	.56
Lectures of reading materials	3.86	.59	4.11	.62
* Practicing listening test questions	3.81	.92	4.25	.69
* Practicing reading test questions	3.89	.88	4.36	.68
* Lectures of listening test questions	3.86	.93	4.28	.56
* Lectures of reading test questions	3.72	1.03	4.39	.64
Lectures of test-taking strategies and skills	3.67	.86	3.78	.92
Regularly-scheduled vocabulary quizzes	4.03	.94	4.14	.68
Completing assigned listening online test questions	3.64	1.07	4.03	.69
Completing assigned listening online test questions	3.67	1.09	4.03	.69
Communicatively-oriented activities ※				
Individual oral/written practice	3.60	.98	3.97	.73
Pair discussion of assigned tasks	3.58	.99	3.94	.92
Group discussion or work	3.57	.92	3.93	.91
Watching English films/completing short film reflections	3.44	.90	3.42	.77
*Sharing of foreign culture and study-abroad experiences	3.67	.92	3.25	.80

※As noted above, the communicatively-oriented activities for TR students included translation and finding test answers for the purposes of test preparation, and were not characterized as knowledge creation or critical thinking.

6 Discussion

This study investigated performance on the GEPT among 72 Taiwanese university students, as well as their perceptions of the effectiveness of traditional and non-traditional test preparation classes in enhancing their English proficiency. Both types of test preparation classes were instructed by the researcher for 54 hours over an 18-week period. The main findings were that (1) the NR subjects achieved greater score gains in listening, while the TR subjects showed greater gains in reading scores; (2) test-unrelated factors, such as limited proficiency in vocabulary, phrase structure, and grammar, were reported by both groups as major concerns in GEPT post-test failure; and (3) traditional teacher-centered activities were reported to be the most effective in enhancing the subjects' English test scores.

While both the TR and NR classes performed significantly better in listening, reading, and overall scores following remediation, the significantly greater gains in reading among TR and among NR students in listening are inconsistent with the findings of Elder and O'Loughlin (2003), Green (2007), and O'Loughlin and Arkoudis (2009), in which initial proficiency positively predicted the degree of score gain. Although the TR subjects' initial proficiency was lower than that of the NR subjects, their score gains in listening were significantly less than those of their counterparts, which implies that initial proficiency may not be the sole factor contributing to such score gains. A possible reason why the NR group showed greater score gains in listening and the TR group in reading may be the different teaching activities in the two groups. The NR group performed more communicatively-oriented activities involving listening tasks, working in pairs or groups to answer reading comprehension questions or holding conversations on a particular topic, whereas the TR group performed more test preparation activities, listening to traditional teacher-dominated lectures, translation of reading materials, and explanation of test questions. These latter activities appear to have enhanced the TR subjects' reading skills.

The subjects' perceptions of the effectiveness of classroom activities were shown to be a salient variable. Teacher-dominated lectures on lexical and grammatical aspects, practice, and the

explanation of test questions were considered by both TR and NR subjects to be more effective in enhancing their English test scores. Teacher-student interaction in oral practice, group/pair practice and discussion of reading materials and test questions, and short film reflections, although considered interesting and practical by the NR subjects, were regarded as less effective than traditional test preparation practices. It may be that progress in communicative competence is more difficult for students to perceive in a short period than vocabulary extension and raised test scores. It appears that changing students' deeply-rooted belief that test preparation courses should familiarize them with the lexical content of the test, train them in test-taking strategies, and give them extensive test practice is a difficult task. However, the large effect size for the TR and NR subjects' perceptions of the effect of teaching activities on test scores, and the TR subjects' low pass rate (1%), indicate that test-oriented instruction may have been a poor motivator for about half of the TR students (see research question 2), not encouraging them to study for the test. This finding may also relate to the TR subjects' positive perceptions of the effectiveness of sharing foreign culture and life experiences in improving their English proficiency. The culture talks of three to five minutes may have offered the TR subjects a break from test-oriented instruction, which they may have found effective but not particularly interesting. This finding is consistent with that of Chu (2009), who reported that the top two types of out-of-school GEPT-preparation practice among 667 Taiwanese university students were entertainment-oriented, namely "listening to English songs" and "watching English TV programs." Similarly, in Alderson and Hamp-Lyons' (1996) study, two ideal test preparation methods were reported to be "having American friends" and "going to the movies." Thus, the design of both interesting and effective test preparation activities for lower-achieving students may require attention. Furthermore, portfolio assessment, including a variety of activities as graduation requirements, may encourage students to develop a broader range of language skills than those that can be assessed by means of a one-off language proficiency test.

The GEPT, adopted as a graduation test by the university targeted in this study, was designed to emphasize communication skills in formal English instruction. However, as their GEPT scores carried significant weight (i.e. 40%) in their graduation, both TR and NR subjects thought that traditional test preparation methods of memorizing vocabulary and grammar rules and practicing tests were more effective in enhancing their English skill than were communicatively-oriented activities. In addition, many of the NR participants wanted vocabulary, grammar, and test practice, and perceived the communicatively-oriented activities as less helpful. This reflects an ethical dilemma for language teachers: teach to the test, because that is what students want (and probably need, in order to pass), or teach them the English they need but that does not necessarily benefit their test performance? A similar situation arises when a test is used for purposes of admission or graduation. For example, local tests, such as the Hong Kong Certificate of Education Exam in English (Cheng, 2004), the National Maturation English Test in China (Qi, 2005, 2007), the College English Test in China (Zhan & Andrew, 2014), and the Basic Competence English Test in Taiwan (Chen, 2006; Lin, 2010), were designed to encourage a shift in English language teaching and learning from a formal linguistic approach to a more communicative approach, but as these tests are for high stakes and play a vital role in determining students' academic futures, teachers devote more attention to assisting students to achieve high scores than to teaching actual communication skills. Similarly, students may prefer traditional methods, such as completing previous tests, drilling, and rote memorization to prepare for the tested skills, and do not adapt their learning methods to a more communicative way of learning. Given the above findings, the mismatch between learner and teacher perceptions regarding the effectiveness of various activities for test preparation, and the mismatch between the test use and the test construct, may lead to unintended washback effects, namely a preference for test-oriented activities, as seen in this study. These findings emphasize the need for collaboration among test stakeholders, such as test developers, policy makers, and educational practitioners, in order to bring about the intended positive effects (Cheng, 2005, 2014).

7 Conclusion

By means of pre- and post-testing with the GEPT, questionnaires, and interviews, this study compared the effectiveness of two types of test preparation classes. Given the differences in score gains between the TR and NR subjects, it remains difficult to determine which type is more beneficial in terms of learning outcomes. The purpose of a test appears to play a vital role in students' preference for test-oriented activities over communicatively-oriented activities for test preparation, although too great an emphasis on test-oriented activities may also be perceived as leading to failures on a post-test. Language testers should note that using a test as a lever to promote intended washback effects, namely improved communicative competence, may not work. Language teachers should also remain aware of their students' preferences and use different strategies to build learner choices into their lessons (Eslami-Rasekh & Valizadeh, 2004). For example, before integrating test-related material into communicative activities, teachers should provide students with sufficient support in terms of fundamental knowledge or practice on vocabulary, grammar, and pronunciation. Whether to teach above or to the test remains an unanswered question, but further investigation may be required to explore the reasons for student preferences for test-oriented activities for test preparation. It is also worth exploring the mismatch, if any, between learner and teacher perceptions regarding the efficacy of test-oriented, communicatively-oriented, and combined approaches to test preparation.

In considering the present findings, certain limitations of this study should be noted. First, the focus was limited to 72 lower-achieving students at one technical university. In order to determine the relevance of the findings to other contexts, future research should include learners of different proficiency levels from a range of learning environments. Second, this study relied on self-reported data and six student interviews. In order to gain a better understanding of the complexity of learner preferences regarding in-class test preparation methods, more in-depth interviews and other types of data, such as learning journals, ought to be employed.

References

- Alderson, C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279.
- Chen, L. (2006). Washback effects on curriculum innovation. *Academic Exchange Quarterly*, 10(2), 206–10.
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 147–170). Mahwah, NJ: Lawrence Erlbaum.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Cheng, L. (2014). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1130–1146). Boston: John Wiley & Sons.
- Chu, H. Y. (2009). *Stakes, needs and washback: An investigation of the English benchmark policy for graduation and EFL education at two technological universities in Taiwan* (Unpublished doctoral dissertation), National Taiwan Normal University, Taiwan.
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test preparation strategies. *Studies in Educational Evaluation*, 45, 62–69.
- Davis, A. (1997). Introduction: The limits of ethics in language testing. *Language Testing*, 14(3), 235–241.
- Elder, C., & O'Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. In R. Tolloh (Ed.), *IELTS Research Report 4* (pp. 207–254). Canberra: IELTS Australia.
- Eslami-Rasekh, Z., & Valizadeh, K. (2004). Learners' voice vs. teachers' voice: Classroom activities viewed from different perspective. *TESL EJ*, 8(3), 1–13.
- Gan, Z. (2009). IELTS preparation course and student IELTS performance: A case study in Hong Kong. *REL C Journal*, 40(1), 23–41.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.

- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches*. Boston: Allyn and Bacon.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport: Praeger.
- Lin, Y. (2010). *A study on the washback effect of the basic competence English test on junior high school students in Northern Taiwan* (Unpublished master's thesis). National Chengchi University, Taiwan.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell.
- Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6(13), 1–30. Retrieved from <http://epaa.asu.edu/epaa/v6n13.html>
- Ministry of Education, (2012). 2006-2009 Annual accountability report (95-98 年度教育部年度績效評估). Retrieved from http://www.edu.tw/secretary/content.aspx?site_content_sn=19757
- O'Loughlin, K., & Arkoudis, S. (2009). Investigating IELTS score gains in higher education. *IELTS Reports*, 10, 95–180.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142–173.
- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education*, 14(1), 51–74.
- Spolsky, B. (1999). Language testing. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics* (pp. 695–703). Oxford: Elsevier.
- 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.
- Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviours* (Unpublished doctoral dissertation). Hong Kong Polytechnic University, Hong Kong, China.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of Medical Education*, 2, 53–55.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28, 499–509.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge: Cambridge University Press.
- Wall, D., & Horak, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 1, the baseline study*. Princeton: Educational Testing Service.
- Wu, J. (2012). GEPT and English language teaching and testing in Taiwan. *Language Testing*, 9(1), 11–25.
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10(2), 196–218.
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self theories. *Assessment in Education: Principles, Policy & Practice*, 21(1), 71–89.

Appendix 1

Teaching activities conducted in NR and TR classes

Teaching activity	NR class	TR class
● lectures of vocabulary instruction	<ul style="list-style-type: none"> ✓ Translation of the key words and phrases in the text/test ✓ Reading and repeating the key words and phrases ✧ Oral questions using the new words/phrases for a review 	<ul style="list-style-type: none"> ✓ Translation of all the words and phrases in the text/test ✓ Reading the repeating all the words and phrases
● lectures of grammar and practice of grammar exercises	<ul style="list-style-type: none"> ✓ Explanation of the key grammatical rules and exercises in the text/test ✓ Presenting the answers to the exercises ✧ Pair dialogue practice focusing on grammatical rules 	<ul style="list-style-type: none"> ✓ Explanation of all the grammatical rules and exercises in the text/test ✓ Presenting the answers to the exercises

● Lectures of reading materials	<ul style="list-style-type: none"> ✓ Translation of the reading materials ✓ Reading and repeating the reading materials ✧ Pair discussion of the questions derived from the reading materials 	<ul style="list-style-type: none"> ✓ Translation of the reading materials ✓ Reading and repeating the reading materials
● Individual oral/written practice	<ul style="list-style-type: none"> ✧ Oral questions using the new words/phrases for a review ✧ Oral questions derived from the reading materials ✧ Completing the study worksheet regarding (1) the learning points, and (2) the difficulties 	<ul style="list-style-type: none"> ✓ Translation practice of reading materials from English to Chinese
● Pair discussion of assigned tasks	<ul style="list-style-type: none"> ✓ Checking the take-home assignments for the mock tests/the text exercises ✧ Pair English discussion of the questions derived from the test materials (eg., What is your favorite attraction that you would introduce to foreigners?) ✧ Pair dialogue practice focusing on grammatical rules 	<ul style="list-style-type: none"> ✓ Pair discussion of the answers to the test questions/text exercises in Chinese
● Group discussion or work	<ul style="list-style-type: none"> ✓ Group discussion of ✧ Compiling a mini-English dictionary using words learned from the text and tests (take-home group project) 	<ul style="list-style-type: none"> ✓ Group discussion of the answers to the test questions/text exercises in Chinese
● Practicing listening/reading test questions	<ul style="list-style-type: none"> ✓ Listening to and answering the listening test questions ✓ Completing the reading test questions 	<ul style="list-style-type: none"> ✓ Listening to and answering the listening test questions ✓ Completing the reading test questions
● Lectures of listening/reading test questions	<ul style="list-style-type: none"> ✧ Asking oral questions related to the pictures/words/topics that appeared on the test (eg., Look at picture 1 and tell me what is next to the tree.) ✓ Presenting the answers to the test questions 	<ul style="list-style-type: none"> ✧ Translation of the listening/reading test questions ✓ Presenting the answers to the test questions
● Lectures of test-taking strategies and skills	<ul style="list-style-type: none"> ✓ Presenting the tips that are often used for EFL listening/reading tests 	<ul style="list-style-type: none"> ✓ Presenting the tips that are often used for EFL listening/reading tests
● Regularly-scheduled vocabulary quizzes	<ul style="list-style-type: none"> ✓ Quizzes about translation of vocabulary from English to Chinese or vice versa 	<ul style="list-style-type: none"> ✓ Quizzes about translation of vocabulary from English to Chinese or vice versa
● Completing assigned listening /reading online test questions	<ul style="list-style-type: none"> ✓ Practicing assigned listening/reading online test questions on the school website 	<ul style="list-style-type: none"> ✓ Practicing assigned listening/reading online test questions on the school website
● Watching English films/completing short film reflections	<ul style="list-style-type: none"> ✓ Watching one English film in class ✧ Oral discussion about the film ✧ Written assignments for the translation of the favorite quotes ✧ Writing answers to short questions about the film ✧ Writing short film reflections in English with some key concepts provided 	<ul style="list-style-type: none"> ✓ Watching two English films in class ✧ Writing short film reflections in Chinese
● Sharing of foreign culture and study-abroad experiences	<ul style="list-style-type: none"> ✓ 3-5 minutes of culture share every week 	<ul style="list-style-type: none"> ✓ 3-5 minutes of culture talk every week