An Exploration into Test-oriented Teaching in a Language Classroom

Hsiu-yu Chu

(Ming Chi University of Technology, Taiwan)

Abstract: There has always been a hot debate on whether the teacher should teach to the test or they should not. The answer varies depending on the situation. This present study explores whether the use of the TOEIC speaking test tasks in Freshman English classes could facilitate students' learning of oral English when most students have recognized the importance of the speaking skill to them. In a Freshman English pre-class survey on students' language needs in a technological university in Taiwan, students overwhelmingly indicated their high interest in oral English than any other language skills. In that regard, three classes of Freshman English in the school were chosen for the present study — two classes, forming the experimental group (EG), were geared towards the improvement of students' speaking ability in spite of the common four-skill teaching materials required for all the freshman students in the school. The other class, the control group (CG), was taught following the common syllabus. In order to emphasize the teaching of speaking to the EG as well as to address the research questions, test-oriented teaching was incorporated into the school's common syllabus for the EG students, focusing on students' mastery of the speaking tasks in the TOEIC test. Quantitative data were then collected from both the EG's and CG's pre- and posttest scores on the TOEIC speaking tests, while qualitative data were gathered through interviews with students in the EG. The results of this mixed-methods study will show whether the test-oriented teaching could improve students' learning of oral English in such a context. Furthermore, the question whether teaching to the test could facilitate or harm learning will also be discussed.

Key words: test-oriented teaching, TOEIC speaking test, Freshman English

1. Introduction

It has long been advocated in the testing literature that testing and teaching should go hand in hand in order to achieve better learning results (Hughes, 1989; Bachman & Palmer, 1996; Bailey, 1996; Wiggins, 1998; Brown, 2001). For example, Messick (1996) stresses the importance of a "logical or evidential link" between the teaching or learning outcomes and the test properties thought to influence them before any possible washback can be claimed. Palmer (2005) proposes a "backward" relationship between teaching and testing, where teachers should at first have their tests in mind and then plan their teaching step by step based on what to be tested after the teaching, instead of teaching first and then designing the test on what they have taught. However, with the intended washback on learning in teachers' mind, it is still not very clear how their test-based teaching can be planned and conducted to trigger the intended washback on students.

Washback refers to "the extent to which the introduction and use of a test influences language teachers and

---

Hsiu-yu Chu, Ph.D., Associate Professor, General Education Center, Ming Chi University of Technology; research areas/interests: language testing, washback and second language acquisition. E-mail: chuhy@mail.mcut.edu.tw.

learners to do things they would not otherwise do that promote or inhibit language learning" (Messick, 1996). There are basically two "directions" of washback: positive or negative effects that a test brings about on teaching and learning. Much has been discussed in the previous studies on promoting positive washback, among which Saif (1999, 2006) proposes that positive washback is likely to be generated if students' language needs are taken into consideration. His study sheds some light on the possible mechanism for washback direction.

Based on the rationale mentioned above, this present study explores the use of the TOEIC speaking test tasks in Freshman English classes, when most students have indicated their needs for better speaking skills and recognized the usefulness of the TOEIC speaking test. In other words, this study aims to test whether the teaching of the TOEIC speaking test could beneficially facilitate students' learning of oral English as well as improve their TOEIC speaking test scores.

2. Literature Review

2.1 Washback on Students

Previous washback studies on students have investigated different aspects of learning such as students' attitudes, feelings, motivations, perceptions, learning strategies and learning outcomes.

Shohamy (1993) in a study on the Arabic Test concludes that students were affected by the test both positively and negatively. What she means by positive washback on learning includes students' learning attitudes, motivations and perceptions. Similar conclusion is also found in Cheng's (1998) study. She investigates the impact of the HKCEE (the Hong Kong Certificate of Education Examination) on Hong Kong secondary school students and finds that the exam did not have a fundamental impact on students' learning, such as their motivation and learning strategies. Cheng argues that it was actually the important role of the HKCEE that worked among students. In Shih's (2007) washback study on the GEPT (the General English Proficiency Test of Taiwan), little washback on learning has been found. Shih studies English majors in two schools in Taiwan, one of which has used the GEPT as the graduation benchmark. He finds that the test brought about only little impact on learning in both schools, though the impact was a little bit higher in the school with the English benchmark for graduation. He argues that, even though the GEPT was widely considered as an important test among students, it had no immediate importance to students.

Among the washback studies on learning, the following studies have particularly focused on learning outcomes, showing inconsistent results.

Andrews et al. (2002) investigate the Hong Kong Advanced Supplementary (AS) "Use of English" (UE) oral exam comparing three cohorts of students and find a very superficial level of learning outcome as reflected by students' familiarization with the exam format and the rote-learning of exam-specific strategies and formulaic phrases. Read and Hayes' (2003) study on the impact of IELTS on preparation for academic study in New Zealand shows about the same result as Andrews et al. (2002). Students' IELTS scores on both the IELTS preparation course and the more general course were found to have increased from pre- to post test. However, the difference in the score increase was not statistically significant either.

Saif's study (2006), on the other hand, shows a more encouraging result on students' learning outcomes. In the study, the ITAs (international teaching assistants) in the experimental group showed significant increases in the scores from pre-test to post test as compared to the control group. However, the study includes a needs analysis before the test development and the test design in the investigation, which possibly contributed to the positive

learning outcomes there.

As reflected in the above literature review, empirical studies on the washback effects on learning are still lacking. The existing studies can provide only little evidence of washback on students' attitudes, feelings, motivations, perceptions, learning strategies and learning outcomes, showing inconsistent results.

2.2 Language Needs in Washback Studies

As mentioned earlier, Saif (2006) proposes that positive washback is likely to be generated if students' language needs are taken into consideration, suggesting a possible mechanism for washback direction. This case study of international teaching assistants (ITAs) is unique in that it includes a needs analysis on relevant stakeholders before developing a new ITA performance test.

Brown (2001) proposes that "needs analysis", defined as "the systematic collection and analysis of all relevant information necessary to satisfy the language needs of the students within the context of the particular institutions involved in the learning/teaching situation", be the first step to designing and maintaining language curriculum, followed by and interacting with developing objectives, writing and using tests, developing materials and teaching, all of which function as the basis of the on-going process of *program evaluation*. Brown's model connects at least two important concepts about positive washback. First, there should be a mutual relationship between teaching and testing (Goldstein, 1989; Shepard et al., 1996; Wiggins, 1998) and second, the test design should appropriately match as many focal constructs as possible (Green, 2007) in order to achieve the intended positive washback. Here the focal construct can be realized as learners' language needs in Brown's model, since needs analysis is usually conducted at the initial stage of curriculum development to determine all the fundamental elements that are to be highlighted at the later stages.

2.3 Research Questions

Two research questions were proposed for the present study:
(1) Could test-oriented teaching improve students' learning of oral English?
(2) How did the use of the TOEIC speaking tasks in class affect students' learning of oral English?

3. Methodology

3.1 Research Context

The research context was situated in a university of technology in northern Taiwan. Only recently the school has divided its freshman and sophomore students into three English proficiency groups (high, medium and low) for the first-year and second-year English required courses.

Student participants were selected into the present study by convenient sampling due to certain limitations for class assignments to teachers in the school. In that regard, two classes of high-level freshman students taught by the researcher were picked as the experimental group (EG). One of the two EG classes was composed of engineering-related majors (EG1) and the other of business majors (EG2), each having around 47–50 students. Another class of students (with engineering-related majors) at the same level taught by another teacher was chosen as the control group (CG). These three classes of students were considered similar to each other as they were recruited into the school and placed at this English proficiency level with the same procedures.

The reason why "high-level" students were chosen for the present study was not just a compromise with the available resources, but there was also a concern about high-level students' language needs. According to Chu

(2009), high-achieving students tended to be more eager to learn English and more clear about what they need English for than low-achieving ones. In that regard, the needs-based approach of the present study would possibly be more productive in the "high-level" classes than in classes at lower levels. Meanwhile, in a pre-class survey on students' language needs in the EG and CG classes, students overwhelmingly indicated their high interest in oral English than any other language skills, and most of them agreed that taking the TOEIC speaking test was a trend in the job market and they would consider taking one in the future.

However, the three-credit Freshman English course in this school used common teaching materials for all the classes at the same level, which focused mainly on general reading and listening skills. The common mid-term and final exams, in a GEPT-like format, were based on the common syllabus for all the teachers teaching the same course at the same level and were administered to all the students at one time during the exam weeks. Therefore, a few teaching adjustments were made in the EG classes to address the research questions.

3.2 Treatments

In order to emphasize the teaching of speaking to the EG as well as to address the research questions, test-oriented teaching was incorporated into the school's common syllabus for the EG students, focusing on students' mastery of the speaking tasks in the TOEIC test.

The official TOEIC speaking test includes 11 questions in six tasks: "reading two short passages aloud" (Q1 and Q2), "describing a photo" (Q3), "responding to three questions about a topic" (Q4, 5, 6), "answering questions using information provided" (Q7, 8, 9), "proposing a solution" (Q10) and "expressing an opinion" (Q11). The instructional tasks were designed using the format of the TOEIC speaking test. There were basically two types of instructional tasks embedded in the EG classes.

(1) **Group work**. There were four tasks that required group work throughout the semester, including "describing a photo", "answering questions using information provided", "proposing a solution" and "expressing an opinion". For each task, a handout was given to students two or three weeks before the presentation week, directing them how to prepare for the given task. Usually the handout included a sample response, useful phrases and expressions for conducting the task, procedures for conducting the task and the assignment itself. The teacher then used the handout to give detailed explanations about each task and demonstrated the sample response in class. In addition to each group's own group presentation in class, students, when sitting as the audience, must be engaged in an activity responding to other groups' presentation. For example, during the group presentation of "answering questions using information provided", each group was required to ask the whole class five or six questions regarding the information they were given beforehand. The information was in the form of a table, diagram or schedule and would appear on student monitors during the Q and A session. After all the groups finished asking their questions, each group got a score for their presentation of questions from the teacher, and they also got a score for the number of questions they answered correctly as audience.

(2) **Individual work**. In order to cover the rest of the TOEIC speaking test — "reading two short passages aloud" and "responding to three questions about a topic" — the teacher often called individual students to read aloud the common reading materials in class, and to answer comprehension questions orally about the reading and listening sections on the textbooks, while going through the common syllabus. Students were awarded individually for their performance on these tasks.

In this way, the common syllabus could still be covered in the EG class for the joint mid-term and final exams for all high-level freshman students, as well as the targeted instructional tasks. Meanwhile, to emphasize

the importance of the TOEIC speaking test tasks in the EG classes, the proportions of students' quizzes, mid-term and final exam were decreased in the EG students' overall grades to give room for the TOEIC speaking test tasks; that is, the four group work tasks and the two individual work tasks in class were also counted towards the EG students' final grades, while the CG students' final grades were decided only by their written quizzes and joint mid-term and final exams.

3.3 Instruments

The instruments used in the study include the following:

3.3.1 Test Instruments

a) pre-test: A TOEIC speaking mock test (one passage from "reading two short passages aloud" was deleted, leaving the total number of questions at 10).

b) post test: Another TOEIC speaking mock test (10 questions as well).

The pre-test and post test were given to the EG and CG in the beginning and at the end of the semester respectively, and the students' answers were recorded in the lab for later grading. One part-time teacher, who did not teach any of the high-level students in the school, was invited to grade all the students' pre- and post test performances. Before formal rating, the rater and the researcher randomly picked ten students, rated the students' recordings together, discussed their differences in scoring and then set the criteria for formal rating. The full score for each test was 100 points.

3.3.2 A Semi-Structured Interview Guide

A list of pre-set questions were prepared to collect data mainly on students' preparation of the TOEIC speaking test tasks in class, their perceptions of the effectiveness of the tasks, the problems they have encountered and their comments on the implementation. A total of ten students — seven from EG1 and three from EG2 — were randomly chosen for the interview. Each was interviewed after class during the later weeks of the semester.

4. Results and Discussion

4.1 Quantitative Results

To answer the first research question about whether test-oriented teaching could improve students' learning of oral English, the paired samples t-test was conducted to compare the pre- and post test scores of EG1, EG2 and the CG respectively. Since there were students absent on the test days, only students who were present on both the pre- and post tests were included in the comparison. Table 1 shows the test results. Although all the three groups made progresses from the pre- to the post test, the difference was only significant for EG2.

Table 1   Paired Samples t-test on EG1, EG2 and the CG between the Pre- and Posttest

					Paired Differences		
Group	Test	n	Mean	sd	t	df	p
EG1	Pre	48	43.71	12.930	.436	47	.665
	Post	48	44.50	16.111			
EG2	Pre	36	51.08	7.549	2.510	35	.017*
	Post	36	54.42	8.597			($p < .05$)
CG	Pre	42	40.76	11.172	1.445	41	.156
	Post	42	42.95	12.445			

4.2 Qualitative Results

The interview data with the ten students in the EG were categorized into the following six themes to answer the second research question about how the use of the TOEIC speaking tasks affected students' learning of oral English:

(1) Preparation for the group work tasks. For the four group work tasks throughout the semester, most interviewees said that they usually did not spend too much time preparing for their presentation (mostly around one hour, before the presentation day). When they prepared for the presentation, a common pattern was that they went on with their own part without discussing the whole thing with other group members (partly because students in one group might not come from the same department and it was not easy for them to meet each other after class). They usually tried to think about what to say in the presentation in Chinese first, and then translated it into English by themselves or by Google and wrote it down. Some more confident students did not even write down the script, but just looked up the unfamiliar English words and phrases in the dictionary. Next they tried to memorize what they wrote or rehearsed what they had in mind before going onto the stage. Only two students in the interview said that they never prepared for the tasks. They just improvised in class.

As mentioned before, students also had to be engaged in related activities responding to the presentation on stage when they were sitting as audience. For this part, most interviewees admitted that it was hard for them to understand their classmates' presentation. As a result, except in the activity, "answering questions using information provided", most audience finished their audience tasks on their own without referring to the presentation made by their classmates in the other three tasks, which did not quite match the teacher's original plan.

(2) Test-oriented teaching and test scores. When asked about how they think the test-oriented teaching could improve their TOEIC speaking test scores, if they were to take one, most students agreed that the use of the TOEIC speaking tasks in class could build up their familiarity with the test format and give them more opportunities to practice, which in turn would increase their test scores. Of all the group work tasks and individual work tasks in class, quite a number of students especially mentioned the expressions taught for "describing a photo", saying these expressions for indicating the locations in a picture (e.g., *at the top/bottom*, *in the lower right/left corner*) greatly helped them to complete the task.

However, one student from EG1 was against the majority view. He believed the improvement of the overall proficiency was much more important than the test-oriented preparation in increasing the test scores. He said the test-oriented preparation could train their "on-the-spot" reaction, but would not necessarily improve the speaking test scores.

In addition, although most students agreed that the more they practice, the better they are supposed to do on the test, some students mentioned that they were not often called to read the text aloud and answer the comprehension questions due to the large number of students in class, which of course reduced their chances of practice on these tasks. As a matter of fact, large-size classes usually have this kind of problems with implementing speaking activities in class.

(3) Test-oriented teaching and overall oral proficiency. When asked about how they think the test-oriented teaching could improve their overall speaking proficiency, most students agreed that the biggest advantage for doing these tasks in terms of their oral proficiency was that they felt they were more confident to speak to the public. In other words, most students think these tasks were only good training for "facing the audience." In

addition to that, very few of them regarded these test-oriented tasks as helpful for improving their overall oral proficiency.

(4) Suggestions for improving test scores. During the interview, students were also asked about their suggestions for improving their TOEIC speaking test scores. Most of them suggested they should practice on old tests, including going on-line for more up-to-date test questions, listening to sample responses and repeating after them, writing down their answers and looking up unfamiliar words and phrases in the dictionary and so on, which basically go with what they were doing for the instructional tasks in class. Only two students (one from EG1, the other from EG2) insisted that the only way to improve their TOEIC speaking test scores is to talk to native-speakers all the time.

(5) Suggestions for improving overall oral proficiency. Students were also asked about their suggestions for improving their overall oral proficiency. A great number of students believed going abroad or talking with foreigners are the best ways to improve their overall oral proficiency. However, three students brought up some ideas other than the dominant view among students: they believed one's overall oral proficiency can be improved by watching English movies and TV series, listening to English songs, role-playing in English in class, practicing speaking English with classmates, making sentences out of new words, surfing the internet in English, reading English magazines, etc. Students seemed to believe that there are more diversified approaches to improving one's overall oral proficiency than to improving one's speaking test scores, and improving speaking test scores alone was not even regarded as one of the approaches to improving their overall oral proficiency.

(6) Students' language needs and the course. In terms of the teaching materials, class activities and examinations, the EG students' language needs were met in different ways. Most interviewees indicated that the textbooks used for this course basically met their language needs, since they contained materials on various topics and provided useful vocabulary and phrases for them. It seems students did not ask too much about the textbooks as long as they offered some "basic stuffs".

As for the class activities, most of the interviewees were reserved about the connection between their language needs and these activities. Most complaints about the instructional tasks were their lack of interactions among people, little chance of individual oral practice in a large class, unfair workload for group work tasks and so on. The only positive view on this part was the extra points they could obtain by correctly answering the teacher's comprehension questions in class; however, this was a little bit far from students' language needs but more associated with the "stakes" of the task to the students.

Most interviewees' attitudes towards the joint written mid-term and final exams were very "traditional". They never thought about the possibility of having alternative types of tests, such as oral tests, to meet their language needs. However, when asked about whether it would be OK to add an oral component to the mid-term or final exam, most interviewees did not quite accept the idea due to their deep worries about their grades.

4.3 Teaching Implications

Some teaching implications from both the quantitative and qualitative results of the study will be discussed in this section especially for language teachers who would like to incorporate test-oriented speaking activities into their classes. The following discussions, though based on this specific research context, might help interested teachers look into some of the possible underlying problems for their context before planning their own teaching and testing.

Influences of test stakes on students' motivations for involvement. Although the quantitative results show

that EG1, EG2 and the CG all made progresses from pre-test to post test, only EG2's progresses were significant. Since the pre- and post tests were only extra tests for all the students, the test scores were not included in the students' formal evaluation. Due to the low stakes of the two tests, students might not have paid too much attention to these tests. Moreover, it is also observable from the interviews with the EG students that most of them did not pay too much effort on either the group work tasks or the individual work tasks in class, probably because these tasks counted for only a small proportion towards their final grades and were not related to the "higher-stakes" mid-term and final exams. In addition, since students were awarded as a group for the group work tasks, some irresponsible students might have got a grade they should not have deserved, which frustrated some hard-working students. It is therefore quite understandable why students ignored these "less profitable" tasks because passing or failing a course feels much more "realistic" to most college students in Taiwan and they probably would rather "invest" their time on other higher-stakes tests. This is especially true to EG1 students, who were exclusively Electrical and Electronic Engineering majors loaded with heavy specialized courses and were often said to haggle over their scores on every subject. Maybe this could explain why such treatments did not have the intended effects on EG1 students.

In sum, to achieve the best results of test-oriented teaching, it is important for the teacher to increase the stakes of the test or task in question to a level that is high enough to trigger students' motivations for involvement.

A closer connection between testing, teaching and language needs. The interview data reveal that most students seem to consider taking a speaking test and improving oral proficiency as two separate jobs; that is, they believed the former could be facilitated through regulated practice but the latter was more complicated and might not be easily achieved through just one single, straight approach. In addition, they tended not to believe higher speaking test scores would guarantee better oral proficiency; on the contrary, they would rather say improving overall oral proficiency could increase the speaking test scores at the same time. To improve overall oral proficiency, students strongly believed in an English-speaking environment, which implies lots and lots of interactions in English in everyday lives. In that regard, it seems the TOEIC speaking test tasks might not be able to reflect students' needs for improving their overall oral proficiency. To better meet students' needs for better speaking skills, it is suggested that teachers do a more detailed analysis on students' Target Language Use (TLU) (Bachman & Palmer, 1996) in order to make an effective teaching and testing plan on the intended TLU tasks.

What is more, teacher-tailored tests, which the teaching would be based on, are believed to be better suited for students' TLU than the TOEIC speaking test, because teacher-made tests could be more flexible and interactional than the one-way TOEIC speaking test. Students are supposed to pay more effort on this kind of tests than simply depend on regulated practice.

Common problems with implementing speaking activities in large-size classes. The qualitative results also brought up one common problem for implementing speaking activities in large-size classes; that is, it is hard for each individual student to have enough speaking practice in class within the limited class time. In this study, quite a number of students complained that they were rarely called to answer the comprehension questions in class. Even in group work, some irresponsible or less motivated students were often reported to have contributed nothing to the whole group.

Another problem teachers might encounter in teaching speaking in large-size classes, as found in this study, is the implementation and scoring of a speaking test. If there is no recording facilities in the lab, the teacher would have to conduct a speaking test on students individually, which is extremely time-consuming. The grading of students' speaking performances, recorded or not, is undoubtedly another laborious job.

Unfortunately, there does not seem to be any better solution to the above mentioned problems. The only way to avoid these problems with implementing speaking activities in large-size classes is to reduce the class size. If small-size classes are impossible for teachers, teachers probably should consider some trade-offs in their teaching and testing plans.

4.4 Limitations of the Study

As mentioned earlier, high-level student participants were selected into this present study due to certain limitations for class assignments to teachers in the school, which might add a constraint to the present study. However, the assumption that high-level students are more prone to needs-based testing and teaching might compensate for such a limitation in a sense. If possible, similar studies could also be conducted on students at different proficiency levels in the future for comparison.

Another limitation lies in the common syllabus and common mid-term/final exams for all the students at the same proficiency level as required by the school policy. Since the common syllabus and the joint mid-term and final exams were based on the school policy, it would not be possible to remove the two exams from the syllabus or digress too much from the syllabus under such circumstances. Neither would it work to largely increase the percentage of the group work or individual work task scores in students' final grades, or even include the two TOEIC speaking tests into the formal evaluation, since the subjective scoring of these tests and tasks could cause fairness problems between the EG and other high-level classes if the intended alignment under the common syllabus was jeopardized.

In general, the school's Freshman English curriculum did impose certain limitations on the present study. From a practical point of view, the teaching-testing renovation, as proposed in the study, would have a better chance of success in a context where not too many constraints were placed on teaching and testing.

5. Conclusion

This study explores the use of the TOEIC speaking test tasks in college Freshman English classes in terms of the effect of teaching to the test, when speaking was considered as the most needed skills among the targeted students and the teaching of it would theoretically bring about positive washback on learning. However, with the certain limitations imposed by the school curriculum on the EG classes, the quantitative results did not quite match the intended outcomes. EG1 students, as well as the CG, did not make significant progresses on the TOEIC speaking test, though both EG1 and EG2 received the test-oriented teaching towards the TOEIC speaking test. The qualitative results show that it was probably because the stakes of the tests and test-based instructional tasks were not high enough to trigger the students' motivations for involvement. It was also observed that the TOEIC speaking test did not fully reflect students' needs for improving overall oral proficiency, which in turn reduced the effect of the test-based teaching. A closer connection between students' language needs, testing and test-based teaching is thus proposed for possibly better outcomes, such as a thorough analysis of students' language needs, teacher-tailored tests which the teaching would be based on and so on and so forth.

Finally, it is suggested that teachers might consider some trade-offs in their teaching and testing plans if they are to conduct similar projects in large-size classes. Undoubtedly, speaking activities generate better results more likely in small-size classes than in large-size classes.

Although the results of this present study did not come along as theoretically intended, it is hoped that the discussions on the problems and difficulties appearing in the study could still shed some lights on the operation of

students' language needs, need-based testing and test-oriented teaching, which is believed to be able to bring about positive washback on learning.

References

Andrews S. (2001). "Reflecting on washback: High stakes tests and curriculum innovation", in: *ILEC Conference*, Hong Kong.

Andrews S., Fullilove J. and Wong Y. (2002). "Targeting washback — A case study", *System*, Vol. 30, pp. 207–223.

Bachman L. F. and Palmer A. S. (1996). *Language Testing in Practice*, Oxford: Oxford University Press.

Bailey K. M. (1996). "Working for washback: A review of the washback concept in language testing", *Language Testing*, Vol. 13, pp. 257–279.

Brown J. D. (2001). *Using Surveys in Language Programs*, Cambridge: Cambridge University Press.

Chen C. (2006). "College students' perception of the impact of graduation English proficiency benchmarks, GEPT and further English learning", in: *Proceedings of the Fifteenth International Symposium on English Teaching*, Taipei, Taiwan: The Crane Publishing Company, pp. 239–249.

Cheng L. (1998). "Impact of a public English examination change on students' perceptions and attitudes towards their English learning", *Studies in Educational Evaluation*, Vol. 24, pp. 279–301.

Cheng L. (2005). "Changing language teaching through language testing: A washback study", *Studies in Language Testing*, Vol. 21, Cambridge: Cambridge University Press.

Cheng L. and Watanabe Y. (with A. Curtis) (Eds.) (2004). *Washback in Language Testing: Research Contexts and Methods*, Maheah, NJ: Lawrence Erlbaum.

Chu H. (2009). "Stakes, needs and washback: An investigation of the English benchmark policy for graduation and EFL education in two technological universities in Taiwan", unpublished doctoral dissertation, National Taiwan Normal University, Taipei, Taiwan.

Goldstein H. (1989). "Psychometric test theory and educational assessment", in: J. Elliott & H. Simon (Eds.), *Rethinking Appraisal and Assessment*, Milton Keynes: Open University Press, pp. 140–148.

Green A. (2007). "IELTS washback in context: Preparation for academic writing in higher education", *Studies in Language Testing*, Vol. 25, Cambridge: Cambridge University Press.

Hawkey R. (2006). "Impact theory and practice: Studies of the IELTS test and Proget to Lingue 2000", *Studies in Language Testing*, Vol. 24, Cambridge: Cambridge University Press.

Hughes A. (1989). *Testing for Language Teachers*, Cambridge: Cambridge University Press.

Hughes A. (1993). *Backwash and TOEFL 2000*, unpublished manuscript, University of Reading.

Hughes A. (2003). *Testing for Language Teachers*, Cambridge: Cambridge University Press.

Messick S. (1996). "Validity and washback in language testing", *Language Testing*, Vol. 13, No. 3, pp. 236–256.

Palmer A. (2005). "Designing authentic test tasks", in: *Proceedings of the Fourteenth International Symposium on English Teaching*, Taipei, Taiwan, pp. 112–121.

Qi L. (2005). "Stakeholders' conflicting aims undermine the washback function of a high-stakes test", *Language Testing*, Vol. 22, No. 2, pp. 142–173.

Read J. and Hayes B. (2003). "The impact of IELTS on preparation for academic study in New Zealand", *IELTS International English Language Testing System Research Reports*, Vol. 4, pp. 153-206.

Saif S. (2006). "Aiming for positive washback: A case study of international teaching assistants", *Language Testing*, Vol. 23, No. 1, pp. 1–34.

Saif S. (1999). "Theoretical and empirical considerations in investigating washback: A study of ESL / EFL learners", unpublished doctoral dissertation, University of Victoria, Canada.

Saville N. and Hawkey R. (2004). "The IELTS impact study: Investigating washback on teaching materials", in: L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in Language Testing: Research Contexts and Methods*, Maheah, NJ: Lawrence Erlbaum, pp.73-96.

Shepard L. A., Flexer R. J., Hiebert E. H., Marion S. F., Mayfield V. and Weston J. J. (1996). "Effects of introducing classroom performance assessments on student learning", *Educational Measurement: Issues and Practice*, Vol. 15, No. 3, pp. 7–18.

Shih C. (2007). "A new washback model of students' learning", *The Canadian Modern Language Review*, Vol. 64, No. 1, pp. 35–162.

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.

Shohamy E. (1993). "The power of test: The impact of language tests on teaching and learning", *NFLC Occasional Papers*, ED 362–040.

Shohamy E. (1997). "Testing methods, testing consequences: Are they ethical? Are they fair?", *Language Testing*, Vol. 14, No. 3, pp. 340–349.

Shohamy E. (2001). *The Power of Tests: A Critical Perspective on the Use of Language Tests*, New York: Longman.

Shohamy E., Donitsa-Schmidt S. and Ferman I. (1996). "Test impact revisited: Washback effect over time", *Language Testing*, Vol. 13, No. 3, pp. 298–317.

Spratt M. (2005). "Washback and the classroom: The implications for teaching and learning of studies of washback from exams", *Language Teaching Research*, Vol. 9, No. 1, pp. 5–29.

Turner C. (2001). "The need for impact studies of L2 performance testing and rating: Identifying areas of potential consequences at all levels of the testing cycle", in: Elder C., Brown A., Iwashita N, Grove E., Hill K. and Lumley T. (Eds.), *Experimenting with Uncertainty: Essays in Honor of Alan Davies*, Cambridge: Cambridge University Press, University of Cambridge Local Examinations Syndicate.

Wall D. (2000). "The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled?", *System*, Vol. 28, pp. 499–509.

Wall D. (2005). "The impact of high-stakes examinations on classroom teaching: A case study using insights form testing and innovation theory", *Studies in Language Testing*, Vol. 22, Cambridge: Cambridge University Press.

Watanabe Y. (2000). "Washback effects of the English section of Japanese entrance examinations on instruction in pre-college level EFL", *Language Testing Update*, Vol. 27, (Summer), pp. 42–47.

Watanabe Y. (2004). "Methodology in washback studies", in: L. Cheng and Y. Watanabe (with A. Curtis) (Eds.), *Washback in Language Testing: Research Contexts and Methods*, Maheah, NJ: Lawrence Erlbaum, pp. 19–36.

Wiggins G. P. (1998). *Educative Assessment*, San Francisco, California: Jossey-Bass Publishers.

Wu J. R. W. (2007). "English language assessment in Taiwan: Where do we go from here?", in: *Proceedings of the 2007 International Conference and Workshop on TEFL & Applied Linguistics*, Taipei: Crane Publishing Company, pp. 574–586.