

Pre- and Posttest Washback in Paired Oral Classroom Assessments

Nathan T. Ducker
Ritsumeikan Asia Pacific
University

Curtis J. Edlin
Ritsumeikan Asia Pacific
University

Richard A. Lee
Kurume Institute of
Technology



Reference Data:

Ducker, N. T., Edlin, C. J., & Lee, R. A. (2013). Pre- and posttest washback in paired oral classroom assessments. In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings*. Tokyo: JALT.

Testing is a vital part of the learning process that teachers and curriculum designers can use to motivate students to study, help them monitor their progress, and guide their pre- and posttest learning activities. Successfully implemented testing should therefore have a positive washback effect on students' learning activities in these areas. To gain full benefit from the testing process, once assessments have been carried out and graded, quality feedback should further help students develop good learning habits and focus their efforts on areas that need attention. This paper reports on the review of a speaking program at a private university in Japan in which the teacher-researchers collected data on the washback effect of a cycle of 8 speaking assessments carried out in one semester, in order to improve the speaking program's efficacy in encouraging learner development through the quality and quantity of pre- and posttest learning practice activities.

学習者のアクションは、言語学習の成功の中心となるが、テストとは、教師やカリキュラム設計者が、学生にやる気を起こさせる為に使用する学習過程で重要なもので、学生が自分の進捗状況を知り、テスト前後の学習を高めるのに役立つものである。それ故、より効果的に実施されたテストでは、学生の学習活動にプラスのウォッシュバック効果をもたらすはずである。テストの過程から最大限の利益を得る為に、課題を実施し採点した後、よりよいフィードバックをする事で、学生はさらに良い学習習慣を生み出し、注意を必要とする分野に努力を集中させる事ができる。この論文は、1学期において8つのスピーキング課題を実施し、質の良い練習課題を数多くこなす事によって、テスト前後の学習者の発達を促しスピーキングプログラムの効果を向上させる為に、教師/研究者がデータを収集した日本のある私立大学のスピーキングプログラムの評価をレポートしたものである。

TESTS ARE tools that, amongst other things, help students develop as language learners (Carr, 2011). Students want to do well within their language courses, and thus tests offer extrinsic motivation with regard to grades (Bernard, 2010). Additionally, learners are also intrinsically motivated by their improvement when the language being tested is meaningful to them (Bernard, 2010). Tests also give students a tangible marker to set goals against, which is an important autonomous learning strategy that leads to better language performance (Oxford & Shearin, 1994). Therefore, when course planners and teachers begin to devise tests and assessments for their classes, it is important that the assessment is judged not only on how reliable and valid it is as a summative tool, but also on its potential to positively impact learning (Black & Wiliam, 1998).

This effect on learning is known as test washback (Alderson & Wall, 1993), and is framed in terms of two dimensions: direction, including positive effects (such as motivating the learner to practice) and negative effects (for example, practicing multiple choice questions at the expense of practicing real language use, or discouraging study altogether), and intensity of washback, referring to either strong or weak effects (see Green, 2007, for a discussion).

The majority of washback research has focused on high-stakes testing while little research has been done on classroom-based testing. One of the reasons that classroom-based assessment may be receiving little attention is the belief that “high-stakes tests have more power to modify teacher and learner behavior whereas low-stakes tests, such as classroom-based assessments, are not central to decision-making and therefore have fewer consequences,” as reported by Munoz and Alvarez (2010, p. 2). However, the need for classroom-based research has been called for by several researchers, (Munoz & Alvarez, 2010; Watanabe, 2005; Xie & Andrews, 2012). Watanabe (2005) argued that more research in this area is needed in order to answer questions such as how to motivate students through tests and to find out what sort of feedback is most useful for students. Furthermore, the majority of the research conducted on washback has dealt with teachers’ responses to tests rather than learners’ reactions with regard to test preparation and follow-up (Xie & Andrews, 2012). Therefore, this study’s goal was to further understand the washback effect of classroom-based testing on students’ learning actions.

The few studies of classroom tests that do exist show that students’ thorough understanding of the expectations and goals of tests plays a large part in determining whether a positive washback effect is produced or not. Munoz and Alvarez (2010) reported that students’ awareness of assessment goals led to them focusing their efforts on better performance on speaking

tests. Similarly, Green (2007) found that students’ understanding of test requirements might be a greater mediator of learning attainment than course content. Additionally, Xie and Andrews (2012) (citing Struyven, Dochy, & Janssens, 2005) suggested that students choose the appropriate learning strategies to match their perceptions of what a test entails. Therefore, it was expected that students’ learning actions would be mainly focused on successful test completion rather than personal language learning goals.

While for some, washback is limited to pretest influence (Peirce, 1992; Berry, 1994), for others, washback has a broader meaning, extending to effects on students taking an exam, feedback received, and subsequent decisions (Bailey, 1999; Brown & Hudson, 2002). Given this wider description of washback, feedback has an important moderating effect on the positive or negative washback of a test. For example, Cameron and Pierce (1994) and Kluger and DeNisi (1996) reported that positively voiced feedback (to encourage students), with no focus on the objective goals of a task, had a negative effect on students’ attitudes toward study and subsequent assessment performance (as cited in Black & Wiliam, 1998, p. 14). The importance of good feedback on students’ successful studying cannot be underestimated, as Hattie (1999) pointed out: “The most powerful single moderator that enhances achievement is feedback” (p. 9). In order to be valuable in terms of positive washback, feedback needs to be diagnostic, detailed, relevant, and useful (Shohamy, 1992; Black & Wiliam, 1998; Munoz & Alvarez, 2010; Munoz, Casals, Gaviria, & Palacio, 2004). Hattie and Timperley (2007) and Black and Wiliam (1998) both further explained that the most effective kinds of feedback involve students both receiving feedback on a performed task and being able to identify how to improve their performance.

In this paper, we report on both the pretest and posttest washback effects of a cycle of speaking assessments conducted eight

times in a semester in a mandatory, intermediate-level, general English course at an international university in Japan. The testing procedure was designed with the intention of maximizing students' speaking opportunities and promoting confidence in their oral abilities. Studying the washback effects of the testing process can help course designers and teachers to understand if a course is well designed in terms of promoting students' proactive, pretest, out-of-class study, and studying the posttest washback effects of the testing cycle can further inform designers and teachers if the test feedback is fulfilling the important educational role of helping students improve their performance.

Course Description and Data Collection

The study was carried out on an intermediate-level, multi-skill, mandatory general English course in an international university in southern Japan with 3,208 domestic (Japanese) students and 2,526 international students from 83 different countries. The majority of students in the course had completed elementary and preintermediate level English classes, while a small proportion of students matriculated directly into the intermediate course on attainment of a paper-based TOEFL score in the 460-479 range. While the majority of students were Japanese, a small number of Korean and Chinese students (fluent in Japanese) studied English alongside their Japanese counterparts and their responses are also included in the data.

The speaking component of the course consisted of eight individual speaking tests developed using task-based role-play activities created from chapter themes and conversation topics contained in the required textbook for the course (Tanka & Most, 2007). The tests emphasized communication strategies. In particular, the main communication strategies were

- initiating conversations,
- introducing topics,

- maintaining conversations,
- overcoming communication breakdown, and
- giving reasons and support.

Students completed the task-based role-plays in pairs, while the teacher assessed task completion and oral proficiency. While two students were completing the assessment, the remaining students carried out other work and waited for their turn. In order to reduce the anxiety associated with testing and to encourage students to feel relaxed during the assessments, the grade for an individual assessment was only 1.5% of the total grade for the course. After the first role-play conversation, all subsequent assessments were designed to recycle and repeat previously covered skills and language using a new topic or context. In this way, each assessment aimed to challenge students to practice previously learned material and reinforce the use of good communication strategies.

The eight assignments were delivered in a cycle of three phases: a task introduction lesson, a task practice lesson, and a task assessment and self-review. In the introduction phase, teachers provided students with a set of worksheets detailing (a) the assessment task and a checklist of the communication strategies upon which teachers' assessments would be based; (b) key vocabulary and language forms useful for satisfactory completion of the tasks; and (c) example conversations (audio file and scripts) and questions designed to raise the students' awareness of the language used by the speakers. For the practice phase, teachers and students were provided with further practice activities and time for students to practice the task with a partner and receive teacher feedback about their general performance. In the assessment phase, students completed a role-play with student partners while the teacher assessed the students' completion of the task using a checklist. Following the assessment, students completed a self-review sheet and teachers gave students feedback related to both their completion of the

task and other areas of their speaking proficiency. Postassessment, students were encouraged to use their feedback to improve areas of speaking proficiency as directed by their teachers; however, no additional class time was set aside for this work. Given the large number of sections (15-20 per semester), it was difficult to determine if the content was consistently delivered in the manner described above.

Testing took place in a very limited time (teachers managed up to a dozen pair interviews in a 95-minute class), and given the complicated nature of the construct of oral proficiency (see Brown, 2003, for a discussion), standardization of grading was difficult. Therefore, to keep the grading uncomplicated and standardized across a large number of sections, the students' assessment scores were calculated based on completion of the task only. Additional feedback was provided on students' oral proficiency, and teachers were encouraged to select one or two areas about which to give students advice on how to improve (see Appendix A). Given the large numbers of sections, teachers, and students involved, it was difficult to determine what feedback was given and how students used it at the time.

Completing three phases of an assessment eight times in one semester was both time and labor intensive for teachers and students alike. However, in a previous study of student activity on the international campus, it was found that despite the setting, students failed to take full advantage of the opportunities to practice English with international students (see Lee, Browne, & Kusumoto, 2011). Therefore, the course designers sought to develop an approach to teaching speaking that would give students as much opportunity as possible to practice speaking in English and further provide students with both the skills to communicate in English on campus and the motivation to practice speaking autonomously. Subsequently, in order to judge the success of the course, we were keen to find out if the testing process promoted students' proactive learning and to what extent.

Additionally, as a testing cycle finished, we wanted to know if students were then able to use teacher feedback to further develop their language practice. If students were not proactively practicing outside of class time and not using their feedback to further improve their proficiency, the designers believed that aspects of the assessment process would need to be redesigned. With this in mind, the following research questions were asked:

- Did the testing process promote students' proactive learning and if so, to what extent?
- As each testing cycle finished, were students then able to use teacher feedback to further develop their language practice?

Data was collected in three stages. A bilingual Japanese / English pilot survey was delivered to two classes, totaling 42 respondents. A follow-up structured interview was carried out with 24 random members of the two classes to check the pilot survey. The questions in the semi-structured interview were initially asked in English and supplemented with Japanese by the interviewer when necessary. The results were recorded on paper, but not digitally. A total of 203 students out of the 327 enrolled in the course responded to the final survey (students who participated in the pilot survey were excluded). The survey was voluntary and anonymous. There was no compulsion for students to take the survey as the course had already been completed. In addition, a bilingual disclaimer explaining the purpose of the survey was included on the first page of the survey.

Results

Unless specifically indicated, responses from the preliminary survey and the interviews reflect the results of the final survey. Key results of the follow-up interview are in Appendix B, and the final survey questions and results are in Appendix C. Once the survey results were collected, the results were analysed for evidence of positive and negative washback.

To answer the first research question (if the testing process promoted students' proactive learning and to what extent), we analysed the responses to Survey Questions 3 and 4 (see Appendix C, Q3 & Q4), which asked about the frequency and duration of student practice. Out of 145 respondents who answered that they practiced, 63% said that they practiced three or more times per testing cycle. Most commonly, students practiced for more than 30 minutes per practice. More specifically, 27.5% of the practices were between 20 and 30 minutes, 35.2% were between 30 minutes and an hour, and 12.4% of students said that their practices were longer than an hour. In terms of their practice foci (see Appendix C, Q10), the majority of respondents (134) reported that practice was aimed at *completion of task*, while the remaining criteria received a nearly evenly distributed numbers of responses: conversational management activities (90), fluency and pronunciation issues (86), accuracy (82), and using the correct vocabulary (84). That students focused on task completion was underscored by the kinds of activities they reported completing in preparation for the test (see Appendix C, Q6). The most popular practice activities were: practicing with a partner from class (84), memorizing key vocabulary (70), planning exactly what to say (68), and writing out a script (67). Further practices with peer advisors, students from other classes, or international students comprised a total of 59 responses.

To further address the first research question, we analysed responses to Survey Question 5 concerning students' motivation for practicing for the tests (see Appendix C, Q5). The majority of students reported that their main motivation was to improve their speaking ability (67.9%), while only 23.6% practiced in order to improve their grade. In fact, the results revealed that only 4% of the students did not practice due to the low weighting of the individual tests.

Additional responses relevant to the first research question were revealed in students' responses regarding the value of the

testing process (see Appendix C, Q15) in that they found the tests to be good for helping them self-monitor their improvement (62%), a good opportunity to converse in English (44.5%), and beneficial in pushing students to study (25.5%). These results slightly contrasted with the interim oral interviews (see Appendix B) in which students indicated that the tests were mainly beneficial in pushing students to study (11 responses), in contrast to student indications that they provided a good opportunity to converse in English (6 responses). Students' positive perceptions of the testing cycle were further highlighted by their high levels of satisfaction with the speaking programme's outcomes (see Appendix C, Q14). In short, students believed that they were improving their oral abilities as they took the tests, which was an integral part of the washback effect of these tests.

To address the second research question (whether students were able to use teacher feedback to further develop their language practice), Survey Questions 8, 11, 12, and 13 were analyzed (see Appendix C). While there may be some variation in how teachers gave feedback, all students were supposed to receive the same grading form from their teacher. Thus, it was important to know if this form could be effectively used by students to review their tests. Multiple items on the grading form were unclear to students. The majority of students were able to discern the meaning of the task's requirements (see Appendix C, Q8). For example, 89.8% said they understood *introduce the topic*, 73.9% said they understood *maintain the conversation*, and 72.6% said they understood *give opinions and support*. However, the linguistic skills pertaining to language proficiency were not well understood. For example, *enunciation* had a positive response of only 39.2%, *syntax* a positive response rate of 46.1%, and *accuracy* a positive response rate of 50%.

In answer to concerns over students' ability to understand the feedback form, 94.5% of students reported that they were able to understand their teacher's written feedback (see Appendix C,

Q11). Yet, the survey results indicate that most students either did nothing with their test results (34.7%), or passively remembered their weak points (46.5%) for the next test (see Appendix C, Q12). Only one student each reported practicing weak points arising from the test results or taking the results to discuss them with a peer advisor; 3.5% reported discussing their results with their classmates; and 6.5% reported discussing their results with their teacher. As for the reasons why students did not review, no single answer clearly stood out as a reason (see Appendix C, Q13). The one result we expected to see more of was *it won't improve my grade*—yet only four students reported this. Conversely, nearly one-fourth of the respondents to this question wanted to review but either did not have sufficient time (21), or did not know how to use their teacher's feedback (25).

Discussion

The testing approach was successful in motivating students to proactively study for the test. Typically, students practiced three times per test for an average of 45 minutes. With eight tests per semester, this results in a typical student completing 18 hours of additional speaking practice—clear evidence of positive washback from the testing cycle. Students usually practiced in at least one of four ways: conversation practice, memorizing vocabulary, writing out a script, or making a list of key points to cover in the test. All of these items focused on the graded portion of the test and revealed that students intended to complete the task and improve their test scores. As no score was given for proficiency items, such as fluency or accuracy, students did not focus on improving their overall oral proficiency. These results correspond with reports that students' learning activities are strongly influenced by perceptions of test requirements (see Green, 2007; Munoz & Alvarez, 2010; Xie & Andrews, 2012). The results allowed us to see that the course achieved two of its

goals by getting students to further practice speaking outside of the classroom and to develop autonomous study habits.

The presurvey interview responses indicated that the tests were a strong motivating factor in making students study and subsequent data collection further supported this. One concern was grade weighting. Considering that it has been argued that low-stakes tests such as classroom-based assessment are not central to decision-making and therefore have few consequences (Munoz and Alvarez, 2010), we were concerned that the points distribution of 1.5% of the students' overall grade would have a negative washback effect on students' motivation to study. However, with only four students responding that the low grade weighting stopped them from studying, the results directly contradicted that particular long-held belief about washback. Additionally, the majority of students reported that their main motivation to study for the tests was to improve their ability to communicate orally in English, rather than to get a good grade. Additional results showing the students' satisfaction with the testing process in relation to communicative ability also support the idea that students perceived the tests as useful in improving their English communication skills.

In contrast to the positive pretest washback of the tests, the posttest effects were mostly negative. An important consideration relating to students' posttest activities was the effect of test design on feedback. Students indicated that they could understand their teacher's feedback; however, many students indicated problems understanding the proficiency section of the grading form. For example, less than 40% of students understood *enunciation*, while 89.8% understood *introduce the topic*. Student responses indicated that teachers were either not taking the time to clarify these words with students or not discussing their impact on students' oral proficiency, which may have been due to the washback effect extending to teachers' actions and their placing more emphasis on the section directly related to

students' scores. This reiterates the need for course designers and teachers to consider how to ensure that feedback given to students is sufficiently diagnostic, detailed, and relevant, as well as understood by the participants, in order to facilitate better use of feedback, as argued by Shohamy (1992), Black and William (1998), Munoz et al. (2004), and Munoz and Alvarez (2010).

The most important finding in terms of posttest washback was that the majority of students did not actively use their teachers' feedback. There could be several reasons for this. Some students reported time constraints—perhaps because there was only a short interval between testing cycles (less than 2 weeks), so there was no time to work on using feedback before the next testing cycle started. Some students cited no additional grades, and a number of students simply had more productive (in their opinion) things to do. Some students reported not knowing how to review; though it was unclear whether this was due to a lack of study skills or a lack of understanding of the technical terms on the grading form. This evidence highlights how important it is that teachers allocate time to help students understand and learn how to use feedback. Finally, attitudes toward tests may impact students' review behaviours. Many Japanese students will have seen previous tests (such as entrance exams) as a barometer of achievement and may not be inclined to see tests as diagnostic tools that carry the requirement of further related study by the student. Whatever the reasons for students' nonuse of feedback, this study highlights that it is the responsibility of course designers and teachers to find ways to actively engage students in well-directed, feedback-driven, postassessment study as suggested by Shohamy (1992), Black and William (1998), and Hattie and Timperley (2007).

Limitations of This Study

Despite a large number of participants, this study should be considered exploratory as it highlights important areas for fur-

ther research in order to fully understand the washback effect of classroom-based tests. The first issue concerns the quality of the data relating to students' practices. We do not know if students had a dual focus when preparing for the tests. For example, when writing out a script, did students aim for higher accuracy or, as they were memorizing vocabulary, did they repeat the vocabulary item many times in a sentence to develop their fluency? This aspect of the quality of a student's practice needs to be further understood to help teachers better advise their students and to help course planners understand the effects the test design has on students' behaviours and subsequent language learning. Additionally, the data indicate that the testing cycle had a motivational effect, encouraging students to study further. Unfortunately, the data are limited; further study in this area would help inform teachers and course designers as to how to better adjust their courses to encourage students to work on their speaking autonomously. Similarly, despite data which indicate that students felt the tests were beneficial in improving their speaking ability, we do not have any data regarding the ways in which this subsequently impacted the washback effect of the assessment cycle. Further investigations into the motivational effects of the testing process would also reveal if students are, after course completion, motivated and able to continue practicing during vacation periods.

Posttest, this study indicates the importance of teacher-assisted focus on feedback. However, we collected no data on the types of feedback that teachers were giving. We need further information on the quality, focus, type, and quantity of feedback in order to discern whether improvements in this area should be focused on course-wide procedures, teacher-centered instruction, or student motivation. Another consideration in terms of postassessment washback and our research design is that many students did use the results of their tests to monitor their progress through test scores. This needs deeper investigation. *Remember my weak points for next time* (see Appendix C, Question

12) can cover a wide range of activities, such as subconsciously readjusting the focus of practice for the following test, actually putting emphasis on checking sentences when writing scripts, or simply thinking, “I hope I get a better grade than last time.” It is difficult for anyone to articulate the mental processes that he or she goes through when describing a learning activity, so future research could employ think-aloud-protocols for data collection to further elucidate the processes that students go through when they prepare and review for tests.

Conclusion

In contrast to other studies on washback, we examined evidence of the pre- and posttest washback effect of an oral classroom assessment cycle by looking at students’ learning actions rather than the effects on teaching. This study has shown that classroom-based oral assessments do have some positive washback effects on learners’ actions before taking a test, as highlighted by students’ further study. The study also indicated that there were negative washback effects, as indicated by the limited range of activities that students undertook. Furthermore, the study suggested that there were posttest washback effects of classroom-based assessments, in that students did not pay attention to feedback and subsequent remedial study. As such, we hope this study provides course planners and teachers with information useful to setting realistic program goals centred on the learner and judging a course’s effectiveness in terms of achieving those goals.

Additionally, given that the evidence pointed to negative washback after a test had been taken, this study highlighted the need for a clear focus on feedback during the assessment cycle. Here again, it is unclear whether teacher action had a mediating role or whether the test procedures and grading led directly to nonuse of feedback. Finally, in order to utilize the potential of classroom-based assessment, not just as a summative tool, but as a practical

way of improving classroom-based language learning, this study highlighted the need for further investigations that consider (a) the actions of learners, such as strategies that students use to manage their practice for tests and monitor progress; (b) data on the motivational processes surrounding tests and students’ perceptions of the testing process; and (c) the strategies that students use once they have received their feedback.

Bio Data

Nathan Ducker is an English language instructor at Ritsumeikan Asia Pacific University, Japan. His research interests include autonomy, speaking skills and the curriculum, and willingness to communicate. He can be contacted at <ducknath@apu.ac.jp>

Richard A. Lee is an English lecturer at the Kurume Institute of Technology, Japan. His research interests are related to speaking skills, curriculum design, pragmatics, study abroad, and the impact of language learning context. He can be contacted at <leera1@mac.com>

Curtis Edlin is an English language instructor at Ritsumeikan Asia Pacific University, Japan. His research interests are largely related to speaking, including prosody, spoken grammars, and the effects of cultural legacy on discourse in speaking. He can be contacted at <edlinc85@apu.ac.jp>

References

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Bailey, K. M. (1999). Washback in language testing. *TOEFL monograph series*. Princeton, NJ: Educational Testing Service.
- Bernard, J. (2010). *Motivation in foreign language learning: The relationship between classroom activities, motivation, and outcomes in a university language-learning environment* (Honors theses, Paper 74). Dietrich College, Pittsburgh, PA. Retrieved from <http://repository.cmu.edu/hsshonors/74>

- Berry, V. (1994). Current assessment issues and practices in Hong Kong: A review. In D. Nunan, R. Berry, & V. Berry (Eds.). *Bringing about change in language education* (pp. 31-34). Hong Kong: The University of Hong Kong, Department of Curriculum Studies.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Brown, J. D. (2003). Promoting fluency in EFL classrooms. *Proceedings of the 2nd Annual JALT Pan-SIG Conference*. Retrieved from <http://jalt.org/pansig/2003/HTML/Brown.htm>
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge, UK: Cambridge University Press.
- Carr, N. T. (2011). *Designing and analysing tests*. Oxford: Oxford University Press.
- Cameron, J., & Pierce, D. P. (1994). Reinforcement, reward, and intrinsic motivation: A meta analysis. *Review of Educational Research*, 64, 363-423.
- Green, A. (2007). *IELTS washback in context*. Cambridge: Cambridge University Press.
- Hattie, J. (1999). *Influences on student learning*. Inaugural lecture: Professor of education, University of Auckland. Available from <http://www.education.auckland.ac.nz/webdav/site/education/shared/hattie/docs/influences-on-student-learning.pdf>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77, 81-112.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Lee, R., Browne, K., & Kusumoto, Y. (2011). Measuring EFL learners' environment: English contact and use outside the classroom at a Japanese international university. *Polyglossia*, 20, 15-25.
- Munoz, A., & Alvarez, M. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27, 33-49.
- Munoz, A., Casals, S., Gaviria, S., & Palacio, M. (2004). Guidelines for oral assessment. *Cuadernos de Investigacion*, Doc. 23.
- Oxford, R., & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *The Modern Language Journal*, 78, 12-28. Retrieved from <http://www.jstor.org/stable/329249>
- Peirce, B. N. (1992). Demystifying the TOEFL reading test. *TESOL Quarterly*, 26, 665-689.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76, 513-521.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30, 325-341.
- Tanka, J., & Most, P. (2007). *Interactions 1: Listening and speaking* (Silver Ed.). New York: McGraw-Hill.
- Xie, Q., & Andrews, S. (2012). Do test design and uses influence test preparation? Testing a model of washback with structural equation modeling. *Language Testing*, 30, 49-70.
- Watanabe, Y. (Interviewee), & Newfields, T. (2005). Insights in language testing [Interview transcript]. *JALT Testing & Evaluation SIG Newsletter*. Retrieved from http://jalt.org/test/wat_new.htm

Appendix A Sample Grading Rubric

Completion of the task: Satisfies the requirements of the test item						
Greet partner (10)	<input type="checkbox"/>					
Introduce topic (10)	<input type="checkbox"/>					
Recommendation 1 (20)	<input type="checkbox"/>					
Recommendation 2 (20)	<input type="checkbox"/>					
Recommendation 3 (20)	<input type="checkbox"/>					
Ask and answer clarification questions (20)	<input type="checkbox"/>					
Score	0-49	50-59	60-69	70-79	80-89	90-100

Hesitations and Halts in Speech, Mispronunciations and Enunciation, Connected Speech	
	◁= Weaker ===== Stronger =▷
hesitations, halts mispronunciations, enunciation connected speech	

Word Order and Agreement, Grammar and Tense, Ability to Comprehend and Negotiate the Dialogue.	
	◁= Weaker ===== Stronger =▷
Syntax (word order and agreement) Accuracy (grammar and verb tense)	

Body Language, Eye Contact, Voice Projection, Ability to Introduce and Maintain Conversation, Overcome Communication Breakdown	
	◁= Weaker ===== Stronger =▷
body language, eye contact, voice projection introducing new topics maintaining the conversation Overcoming communication breakdown	

Range and use of vocabulary is content appropriate, English only (i.e., no unnecessary use of first language)	
	◁= Weaker ===== Stronger =▷
vocabulary range	

Appendix B Selected Responses from the Structured Interview Questions

Do you think that these tests are useful for improving your English?

	Response count
Yes, the content is useful	4
Yes, the tests make me study / practice	11
Yes, I can understand how to improve my skills	1
Yes, they are a good chance for us to really speak English	6

We have a lot of tests on this course - do you think this number is too many, just right, not enough?

	Response count
Too many	8
Just right	13
We should do more	3
No response	2

Appendix C

Finalized Survey and Results (irrelevant data tables omitted)

3. How many times do you usually practice as homework for a speaking test? 宿題として、大体何度スピーキングテストの練習をしますか?

Five times 5回	Four times 4回	Three times 3回	Two times 2回	One time 1回	Never 全くしない
15.9%	10.3%	37.2%	28.2%	9%	2.8%

Note. $n = 145$.

4. If you practice, how much time do you spend practicing? 練習にどれくらいの時間を費やしますか

More than 1 hour 1時間以上	More than 30 minutes 30分以上	20 - 30 minutes 20-30分	10 - 20 minutes 10-20分	0 - 10 minutes 10分以下
12.4%	35.2%	27.5%	21.4%	3.4%

Note. $n = 145$.

5. I prepared carefully because 私は熱心に準備しました、何故なら...

I worried about getting a high score for my GPA GPA 高得点を取り、GPAを上げたいから。	I always prepare carefully for tests 私は常にテスト勉強を熱心にするから。	I wanted to improve my ability to speak スピーキング能力を向上させたいから。
23.6%	8.6%	67.9%

6. If you do practice at home, which of these activities do you do to practice? 下記のどのような方法で練習しますか?

	Response count
Listening to the audio files on blackboard ブラックボードの音声ファイルを使う。	21
Practicing with a partner from class 同じクラスのパートナーと一緒に練習する。	84
Practicing with a partner from another class 他のクラスのパートナーと一緒に練習する	22
Practice with an international student 国際学生と一緒に練習する。	26
Practice with a PA* from SALC* SALCのPAと一緒に練習する。	11
Write out a script 台本を書き出す。	67
Memorize key vocabulary 重要な単語を記憶する。	70
Practice key phrases 成句を練習する。	29
Practice the key grammar 重要な文法を練習する。	34
Use shadowing シャドーイング(聞いた英語をすぐに追いかけて声に出す学習法)する。	21
Plan exactly what to say 何を言うか全て決めておく。	68
Other その他(詳しく書いて下さい)	2

* PA - Peer Advisor (formerly called Teaching Assistant)

**SALC - Self Access Learning Center

7. Which of these is true for you:

下記のどれが当てはまりますか

I didn't prepare because

私はあまり準備をしませんでした、何故なら・・・

	Response count
The grade was only 1.5 % of my total grade テストの評価は全体の1.5%でしかないから。	4
The tests didn't motivate me やる気が出る課題がないから。	5
I had other more important things to do 他に優先すべきものがあるから。	22
I didn't know how to prepare 準備の仕方が分からないから。	9
I had prepared enough in class 授業中に与えられる時間だけで十分だから。	13

8. Which of these words from the speaking test form do you understand? スコアシートに記されている、どの項目を理解していますか?***

	Yes, I understand 完全に理解している項目	I am not sure どちらともいえない	No, I don't understand 全く理解していない項目
Introduce the topic	89.8%	10.2%	0.0%
Use transitions to signal questions	57.4%	36.1%	6.5%
Maintain the conversation	73.9%	22.5%	3.6%
Give opinions and support	72.6%	22.6%	4.7%

	Yes, I understand 完全に理解している項目	I am not sure どちらともいえない	No, I don't understand 全く理解していない項目
Close the conversation	76.6%	19.6%	3.7%
Hesitations	54.6%	31.6%	13.8%
Halts	39.8%	37.5%	22.7%
Mispronunciations	56.1%	29.2%	14.6%
Enunciation	39.2%	35.7%	25.1%
Connected speech	69.2%	23.8%	7.0%
Syntax	46.1%	35.4%	18.5%
Accuracy	50.0%	31.6%	18.4%
Body language	87.2%	11.7%	1.1%
Voice projection	79.0%	17.0%	4.0%
Introducing new topics	82.3%	15.4%	2.3%
Maintain the conversation	84.1%	13.5%	2.4%
Overcoming communication breakdown	62.4%	28.8%	8.8%
Vocabulary range	78.0%	19.7%	2.3%

***A Japanese translation of these items was not included at this stage in order to determine if students understood the English only grading rubric and English only teacher explanations

10. If you practice, what do you focus on? (You can choose more than one). 練習する際、何に注意しますか?(複数選択可)

	Response count
I didn't practice 私は準備をしませんでした。	15
Completion of task / Completing the conversation (introduction, opinions, questions, maintain the conversation, closing) 課題にそって会話を発展させること。(前置き、意見、質問、会話を保つ、結句)	134
Hesitations, halts, pronunciation, enunciation, connected speech ためらい、口ごもり、発音あやまり、発声	86
Word order, subject verb agreement, grammar, tense 語順、呼応、文法、時制	82
Body language 身振り、手振り	49
Introducing, maintaining conversations, overcoming communication breakdown 説明、会話を保つ、途切れた会話からの立ち直り、	90
Vocabulary 語彙	84
I don't choose one thing, I just try to complete the conversation 会話をとぎれさせない練習はするが課題内容チェックはしません。	11
Other その他(詳しく書いて下さい)	5

Comprehension of feedback

11. Which is true for you? 下記のどれが当てはまりますか?

I understand my teacher's written feedback	I don't understand my teacher's written feedback	I can't read my teacher's written feedback
先生が書いたフィードバックの内容を理解出来ます。	先生が書いたフィードバックの内容を理解出来ません。	先生が書いたフィードバックが読めません。
94.5%	3.0%	2.5%

12. What do you usually do with your feedback form? フィードバックの内容を見てあなたはどのようにしますか?

I look once to check my score, but I don't review 自分の得点は確認しますが復習はしません。	34.7%
I do nothing 特に何もしません。	7.9%
I look and remember my weak points for next time フィードバックを満見で次回の為に自分の弱点を覚えておきます。	46.5%
I review my teacher's written feedback and discuss with my teacher 先生が書いたフィードバックを元に復習し、先生に助言を求めます。	6.4%
I review my teacher's written feedback and discuss with a SALC PA 先生が書いたフィードバックを元に復習し、SALCのPAに助言を求めます。	0.5%
I review my teacher's written feedback and discuss with my classmates 先生が書いたフィードバックを元に復習し、クラスメイトに助言を求めます。	3.5%
I look at the form, then I practice my weak points carefully フィードバックを見て自分の弱点をよく練習します。	0.5%

If you practice after the test please explain how

テストの後に練習した事があれば、どのように練習したか教えて下さい。

13. If you don't review, please can you explain why:
 テストの後フィードバックを参考にしない理由は次のうちどれですか？

	Response count
The teacher didn't tell me to review 復習する様に言われてないから。	5
I don't have enough time 時間がないから。	21
It won't improve my grade 成績に関係がないから。	4
I don't know how to review 復習の仕方が分からないから。	25
I am not interested 興味が無いから。	7
I had other more important things to do 他にやるべき事があるから。	17
I had more fun things to do 他に楽しめる事があるから。	3
Something else? その他。出来るだけ詳細に説明して下さい。	2

14. Considering the speaking test, which of these things do you think you have specifically improved this semester? スピーキングテストを考慮した上で、下記のどの項目が特に上達したか教えて下さい。

	I have definitely improved this 間違いなく上達した項目は	I have may-be improved this 上達したかもしれない項目は	I have not improved this 上達していない項目は
Choosing correct vocabulary in conversations 会話の中で正しい単語を使う能力	31.4% (61)	59.3% (115)	9.3% (18)
Using correct grammar in conversations 会話の中で正しい文法を使う能力	32.1% (63)	54.1% (106)	13.8% (27)

	I have definitely improved this 間違いなく上達した項目は	I have may-be improved this 上達したかもしれない項目は	I have not improved this 上達していない項目は
Speaking smoothly – (fluently) スムーズに話す能力	52.0% (102)	39.3% (77)	8.7% (17)
Speaking quickly – (fluently) 早く話す能力	40.4% (78)	46.1% (89)	13.5% (26)
Correct pronunciation 正確な発音能力	29.2% (56)	54.7% (105)	16.1% (31)
Correct intonation 正確なイントネーション	30.1% (58)	48.7% (94)	21.2% (41)
Confidence in speaking 自信を持って話す能力	52.8% (102)	37.3% (72)	9.8% (19)
Speaking skills (such as starting a conversation with a stranger or explaining again if your partner doesn't understand) 会話能力(例—他人と会話を始められる)(英語で自分から話しかけるスキル)	49.5% (96)	42.8% (83)	7.7% (15)
Speaking on more complicated topics than before 以前よりも難しい話題について話す能力	42.3% (83)	43.4% (85)	14.3% (28)
Speaking with international students better than before 以前よりも国際学生と上手く話す能力	45.1% (88)	43.1% (84)	11.8% (23)

	I have definitely improved this 間違いなく上達した項目は	I have maybe improved this 上達したかもしれない項目は	I have not improved this 上達していない項目は
Talking about a wider variety of subjects than before 以前よりも広域の話題について話す能力	38.7% (75)	47.9% (93)	13.4% (26)

15. Do you think that speaking tests are a good way to improve your English? (You can choose more than one). 英語の能力を高める為にスピーキングテストは役に立つと思いますか？(複数選択可)

Yes, I can check my improvement はい。英語力の上達が確認出来ます。	62% (124)
Yes, they make me study はい。テストが勉強する動機になります。	25.5% (51)
Yes, I have a chance to speak English 英語を話す機会が持てます。	44.5% (89)
No. Can you explain?	1.5% (3)

16. We have had 8 speaking tests this semester. Do you think this number is... 今回のセメスターで8回のスピーキングテストを行いました。この回数についてあなたはどのように思いますか？

too many 一多すぎると思います。	23.3% (47)
just enough 一ちょうど良いと思います。	65.8% (133)
not enough 一十分ではないと思います	5.9% (12)
I have no opinion 一特に意見がありません。	5.0% (10)