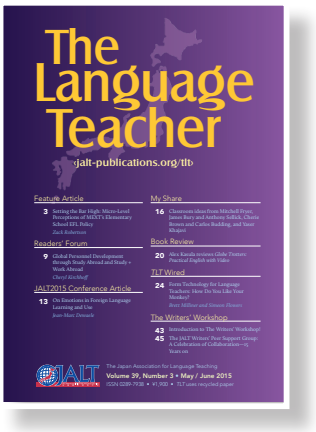
JALT Publications • Online Journals

The Language Teacher



The Language Teacher (*TLT*) is the bimonthly publication of the Japan Association for Language Teaching (JALT). It publishes articles and other material related to language teaching, particularly in an Asian context. *TLT* also serves the important role of publicizing information about the organization and its many events.. As a nonprofit organization dedicated to promoting excellence in language learning, teaching, and research, JALT has a rich tradition of publishing relevant material in its many publications.

Links

- JALT Publications: http://jalt-publications.org
- *The Language Teacher*: http://jalt-publications.org/tlt
- *JALT Journal*: http://jalt-publications.org/jj
- *Conference Proceedings*: http://jalt-publications.org/proceedings

- JALT National: http://jalt.org
- Membership: http://jalt.org/main/membership

Provided for non-commercial research and education.
Not for reproduction, distribution, or commercial use.

JALT
THE JAPAN ASSOCIATION FOR LANGUAGE TEACHING
全国語学教育学会

Building Success: Task-Based Speaking Tests in the Japanese Classroom

Roy Morris
University of Wollongong

This paper outlines the use of formative and task-based speaking assessments in the Japanese classroom at a range of levels, and argues that integrating this kind of assessment into a program benefits students by building their confidence through progressive achievement and by providing many chances for form negotiation and washback. An example of such a speaking test is outlined, and then evaluated according to criteria outlined by Brown and Abeywickrama (2010), and in accordance with the theories of Nation (1996), Long and Porter (1985) Swain and Lapkin (1989, 1998), and Ellis (1997), amongst others. The paper concludes that task-based speaking assessments help to encourage students to speak English by not only providing opportunities to do so, but also by clearly communicating successes and progressive milestones to them.

本論は、日本の教育現場での形成的かつタスクに基づいたスピーキング評価の使用について概説したものである。その評価方法を教育課程に組み入れることによって、学習者は多くの形式交渉や、それらの波及効果による段階的な習得を通じて、言語能力に自信を持つことができると本論で論じる。Brown & Abeywickrama (2010) の評価基準に基づき、また Nation (1995)、Long & Porter (1985)、 Swain & Lapkin (1989, 1998)、Ellis (1997)等の理論に従い、1つのスピーキングテストを例として挙げる。本論の結びとして、タスク中心のスピーキング評価は、学習者に単に英語を話す機会を与えることだけでなく、彼らに段階的な達成目標を明確に認知させ、成功体験をさせることによって、英語を話すことに意欲をもたせることができると結論づける。

This paper outlines and argues for the use of task-based speaking tests as a form of formative evaluation, and in particular as a tool of encouragement and motivation amongst lower level students. The tests are used in a variety of ways, and can provide a facet of a broader spectrum of evaluation, alongside tests designed to assess reading, writing, listening, and presentation skills. The paper will outline an example of a speaking test that I have used which is suitable for a Japanese senior high class, and which provides opportunities to negotiate form and meaning (Swain, 1995) both with the teacher, and with each other (Ellis, 1997; Fernández-García & Martínez-Arbelaiz, 2002; Long & Porter, 1985; Swain & Lapkin, 1998). I will refer to guidelines for assessment outlined by Brown and Abeywickrama (2010).

Background

Free speaking in class is often the first part of lessons to be cut due to time restraints in Japanese high schools, and subsequent speaking activities tend to be restricted to confirming answers of grammar questions, choral speaking, and closed-form interrogational questions as Long and Porter discuss (1985, p. 209), while speaking assessment in particular can dwindle to almost nothing. This is a serious imbalance, and only aids to reinforce in students the idea of English as a theoretical language useful only for passing tests. Speaking tests not only can measure spoken English competency directly, but also, they can act as a motivator, by demonstrating to students their own competence in speaking ability and providing evidence of practical applications of English.

Although many students in Japanese educational systems do learn English at a level that satisfies the requirements of their exams, there are many students with paradoxically low ability who never seem to improve, or even some who are repelled by English. While this can happen when a course is mandatory for all students, this attitude may be bolstered by the relative success of their peers and the lack of general feelings of success they themselves experience. Another thing to consider is different learner styles and needs. Andreou, Andreou and Vlachos (2008) show that students with different learning styles need different kinds of input and output in the classroom; in this way assessment should also be varied in order to not discourage or discriminate.

Criteria

Brown and Abeywickrama's (2010) criteria for assessing the quality of assessment, which I will use here, holds true for both formative and summative tests, and can be very useful when analyzing and assessing the tests you make and use. They outline the five major principles involved in creating and administering language assessment as practicality, reliability, validity, authenticity, and washback. While a thorough reading of the text is best, I will summarize the points here.

Practical tests, simply put, do not overly tax budgetary or time constraints set upon a teacher or student. As Brown and Abeywickrama state, "In classroom-based testing, time is almost always a crucial practicality factor for busy teachers with too few hours in the day" (2010, p. 27).

Reliability is concerned with the accuracy at which a test can reflect the true ability of a student. Issues like repeatability of a test or assessment are paramount. Do students who do the test at different times get a different score due to distracting factors or poorly prepared materials? Issues of inter- and intra-rater reliability are also important to consider (Brown & Abeywickrama, 2010, p. 41).

The validity of a test is concerned with content and impact—in other words, does the test logically follow on from classwork, and indeed measure what it is testing for? Is the test well prepared for, and does it offer opportunities for learning? Does it present itself as a punishment for students, or an encouragement to "bring out the best in their performance" (Brown & Abeywickrama, 2010, p. 44)?

Another important and problematic facet of validity that needs to be addressed in Japanese schools in particular is face validity—"…the degree to which a test *looks* right, and *appears* to measure the knowledge and abilities it claims to measure, based on the subjective judgment of the examinees who take it" (Mousavi, 2009, p. 247, as cited in Brown & Abeywickrama, 2010, p. 35).

An authentic test or assessment usually has some practical real-world application or relevancy, contextualizing language rather than isolating it. Issues with authenticity come about when language is isolated, forced, unnatural, or irrelevant to the learner (Brown & Abeywickrama, 2010, p. 38). Tests should provide the freedom to use non-target language to achieve the same goals.

Finally, washback refers to the ability of a test to effect a change in the behavior of teachers or students, and to encourage them to try things "they would not necessarily otherwise do" (Alderson & Wall, 1993, p. 117). This should be a prime concern in any classroom assessment. Assessing students can have positive or negative results on their future learning. Good tests should positively affect both students and teachers, offer preparation time, give learners constructive criticism, and be ongoing; that is, formative in nature.

Example of Speaking Assessment

Here is an example of a practical speaking test that follows these guidelines.

Students either work alone or in pairs to create the short script of a conversation or short presentation about a prescribed topic—for example, a prompt might be given that reads "Which do you think is better, travelling by train or by car?" The monologue or dialogue is then written down—usually, it is about 40–50 words, or four turns taken while talking—and checked by a teacher, who provides immediate feedback for students about length, structure, grammar and spelling, among others. Teachers can correct the speech outright, actively negotiating with students to reach the final form; alternatively they can underline or highlight problem areas, giving hints as to the type of mistake, and requiring students to correct their own language or to find help from other students or textbooks.

After the presentation is corrected and polished, students memorize it before returning to the teacher and presenting it. This is quantitatively assessed, not qualitatively—students finish as many of these tests as possible in the term, and are given a score based on the amount of work they do, not the quality of the presentations themselves. Teachers can control for quality by not allowing conversations which are too short or simple to be assessed.

Analysis of the Test Procedure and Rationale

The purpose of this speaking test is to encourage meaning-focused output to complete the task. This goal adheres to Swain's Output Hypothesis (1995), which looks at output as having three functions: noticing the gap, hypothesis testing, and a metalinguistic function, that is, negotiation of form. These goals are in line with sociocultural theories put forth by Long and Porter (1985), Appel and Lantolf (1994), and Wertsch (1991). However, how does it measure up as assessment according to Brown and Abeywickrama (2010)?

On the face of it, the issue of practicality seems glaring: how can a teacher expect to simultaneously supervise and control a classroom of students, and administer a test to one or two students at a time? Of course, this type of test would be more easily administered in a team-teaching situation, as my situation was. Technology can also help solo teachers to better utilize class time: presentations or conversations in the second example can easily be recorded in class with minimal teacher supervision, and allows students the freedom to perform when they feel ready. Alternatively, students could record themselves outside of the classroom and submit it to the teacher at their convenience, an activity similar to what Nation calls "the best recording" (1991, p. 4) in which students repeatedly record their

voices in order to fix their own mistakes. Though this admittedly suffers from reliability issues—some students may be too busy to spare the time, and in areas with disadvantaged students, recording devices may have to be supplied. This could encourage study outside the classroom—something which Fukuda and Yoshida (2013) discuss as a substantial problem in Japanese education.

Questions of intra-rater reliability can cause concern in some teaching situations if the classes are shared between teachers. It is partly for this reason that these tests are fundamentally quantitative, not qualitative. A significant side benefit of either test is found in letting the students be creative in what they output. The other reason for stressing quantity over quality is that these tests are designed to encourage what Nation describes as essential strands of a language course: "Learning through meaning-focused output" and "becoming fluent with what is already known" (Nation, 1996, p. 7).

There are some problems with reliability and repeatability in that some teachers accept lower levels of English output as adequate for progression than other teachers; therefore a strong and clear rubric is required for these tests, in order to help standardize the tests as much as possible across a teaching staff. As the tests are criterion-referenced and formative, a bare minimum is set, and ample encouragement to exceed it is given by teachers. We are trying to foster growth on a personal level, and would disadvantage students by setting global standards (Brown, 1995).

Content validity is addressed by matching the assessment to other coursework. A sample criterion might read "answers should include one present perfect sentence", or students may be asked to disagree with each other or concede the point during the conversation. In this way, the final decision about what to say rests with the students, but some form of guidance can also be administered. Another measure of validity is in its direct testing of spoken English. Although "'direct assessment' is a misnomer because it always promises too much" (Messick, 1996, p. 244), it is still more valid than the much less direct method of assessing speaking through listening tests sometimes relied upon.

Face validity, on the other hand, can be a problem. Formative speaking tests, when used within a system whose dominant paradigm is formal summative exams, can be seen as challenging the norm. However, this is not entirely a negative: many students hate and fear summative exams, as Brown and Abeywickrama (2010, p. 1) discuss: "The fear of failure is perhaps one of the strongest negative emotions a student can experience, and the most common instrument inflicting such fear is the test." A simple semantic shift—calling the test a *checkpoint*, for example—can benefit both the teacher and student, shifting the focus from the fear of failure that a *test* might engender, and reframing the event as a measure of relative success, while still borrowing from the semantic meaning, and thus instilling a weight of importance to the activity. In my experience, classroom participation and motivation have risen since the implementation of the test.

The speaking test does not deal directly with the issue of authenticity, and is indeed rather inauthentic in its application, asking students to engage in stilted turn-taking that bears little resemblance to authentic conversation. However, the aim is to encourage critical thinking and opinion making, not to mimic real discussion. The language features teachers test for may in some ways dictate this. If one were to make a speaking test where, for example, one student assumed the role of a used car salesman, and another had to negotiate a good price for a hypothetical car, the thematic arrangement of the test could provide opportunities to use more authentic English.

Perhaps the way it is most useful is in its immediacy of feedback. In a school system there are few opportunities for one-on-one work; however during this test, students can test hypotheses and try new English without fear of failure. There is immediate washback for students and teachers alike during the tests, in seeing which grammar points are troublesome, which language is popularly used, and which students need extra help. The benefit of personalizing work like this is that students get immediate personalized attention and tuition.

Summation

In my experience using this test, students have been free to test their own hypotheses about the language and to think of English as a practical tool of communication rather than isolated phrases learned by rote. Working in pairs has the added bonus of encouraging them to negotiate form (Swain, 1995). Further, automatization of the language (Nation, 1996) is encouraged through repetition of the exercises.

Lastly, it is important to show students evidence of their own progress, not merely their competence, if we are to help them feel empowered and motivated in their study. When I first used the speaking test, many students couldn't manage their time effectively and rushed at the end; however I found in repeated usage that their time management skills also improved. Many of my worst students in fact

seemed to enjoy the assessment, using it as a way to compete against others in a friendly way. I would call that a confidence-boosting success.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*, 115-129.

Andreou, E., Andreou, G., & Vlachos, F. (2008). Learning Styles and Performance in Second Language Tasks. *TESOL Quarterly, 42* (4), 665-674.

Appel, G., & Lantolf, J. P. (1994). Speaking as mediation: A study of L1 and L2 text recall tasks. *The Modern Language Journal, 78*(4), 437-452.

Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment, Principles and Classroom Practices* (2nd ed.). New York: Pearson Education.

Brown, J. (1995). Differences between norm-referenced and criterion-referenced tests. In Brown, J. & Yamashita, S. (eds.), *Language Testing in Japan*, Japan Association for Language Teaching, Tokyo, 12-19.

Ellis, R. (1997). *SLA research and language teaching.* Oxford, UK: Oxford University Press.

Fernández-García, M., & Martínez-Arbelaiz, A. (2002). Negotiation of Meaning in Nonnative Speaker-Nonnative Speaker Synchronous Discussions. *CALICO Journal, 19*(2), 279-294.

Fukuda, S., & Yoshida, H. (2013). Time is of the essence: factors encouraging out-of-class study time. *ELT Journal,* 67(1), January, 31-40.

Long, M., & Porter, P. (1985). Group Work, Interlanguage Talk, and Second Language Acquisition. *TESOL Quarterly, 19*(2), 207-228.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*, 241-256.

Nation, P. (1991). Fluency and Learning. *The English Teacher 20*, 1-8.

Nation, P. (1996). The Four Strands of a Language Course. *TESOL in Context*, *6*(2), June, 7-12.

Richards, J., & Rodgers, T. (2001). *Approaches and methods in language teaching* (2nd ed). New York: Cambridge University Press.

Swain, M. (1995). Three Functions of Output in Second Language Learning. *Principle & practice in applied linguistics: studies in honour of H.G. Widdowson,* Guy Cook and Barbara Seidlhofer (eds.), Oxford, Oxford University Press, 125-144.

Swain, M., & Lapkin, S. (1989). Canadian Immersion and Adult Second Language Teaching: What's the Connection? *The Modern Language Journal*, *73*(2), 150-159.

Swain, M., and Lapkin, S. (1998). Interaction and Second Language Learning: Two Adolescent French Immersion Students Working Together, *The Modern Language Journal, 82*(3), Special Issue: The Role of Input and Interaction in Second Language Acquisition (Autumn, 1998), 320-337.

Wertsch J. (1991). *Voices of the mind: A Sociocultural approach to mediated action.* Cambridge, MA: Harvard University Press.

Roy Morris has taught English in Japan in a range of situations, from early elementary education through to mature-aged students, though most of his experience involves upper secondary and tertiary students. He recently graduated with distinction from the University of Wollongong in NSW, Australia. His focuses of study were speaking and assessment. He currently works in a technical college specializing in English education in Kanagawa, Japan.

2015 CUE Conference

JALT College and University Educators Special Interest Group

Sugiyama University, Nagoya September 26-27

Diversity: Communication and Education

Featured speakers: Tadashi Shiozawa & James D'Angelo

Call for proposals closes July 15

conference.jaltcue.org/call

Image: CC2 Flickr user CANNIK