
**IDEA SHARING: PRACTICING AND CONDUCTING ORAL
TESTS IN THE CLASSROOM**

David Evans

The National College of Nursing

John Herbert

Hiroshima Jogakuin University

Introduction

There are many directions that English language teaching approaches have taken in the past, including the grammar translation method, direct method, audio-lingual approach, and communicative teaching (Crystal, 2007, pp.437-440). Following on from the communicative approach that became popular in the 1970s, we have seen a move towards content-based teaching, and content and language integrated learning (CLIL). Moreover, as the concepts of English as an international language (EIL) and English as a lingua franca (ELF) grow in importance, the debate over which kind of English we should be teaching continues. In spite of the changing English teaching context, in many institutions in Asia, and indeed the world, there is still a place for classes named *English Conversation*, *English Communication*, or *Oral English*. This is not surprising considering that the ultimate goal of many educators in the English teaching world is to produce students who become good speakers of English.

The focus of this paper is the English conversation class, and why and how to use oral tests to evaluate students in these classes. After setting out the rationale for conducting oral tests, this paper outlines a model for a conversation test and a peer evaluation

procedure that help develop student autonomy. Finally, the paper will discuss the advantages and disadvantages of such an approach. This model is based on the authors' many years of experience of using oral tests with non-English major students in Japan. The paper is, therefore, culturally biased (or at least nuanced) in that the conversation classes and oral tests referred to are based in Japan. However, many of the principles and ideas in this paper can be applied to different English language classroom contexts in other parts of the world.

The Japanese context

In Japan, the stereotype of an English learner is of a person who is shy, afraid and ashamed of making mistakes, and often answers questions with the minimum number of words possible (See McVeigh (2002) who addresses this issue in relation to the Japanese higher education system). The following is an illustrative example of a typical exchange in a Japanese college class, where the teacher has to drive the conversation by supplying extra questions, due to the short answers supplied by the student.

Teacher: What did you do in the summer vacation?

Student: I went to America.

Teacher: Where did you go?

Student: San Francisco.

Teacher: What did you do while you were there?

Student: I saw a baseball game.

No matter how strongly we confirm that this is simply a stereotype, the description will have the ring of truth to many English language educators in Japan. The stereotype of Japanese being poor English users is something that the Japanese government is trying to eliminate. English was first introduced as an elective subject in secondary education (in junior high schools) in Japan in 1947, but it was not until 2002 that it actually became a required subject for the

six years of secondary school study (Gottlieb, 2005, p.31). In recent years, in order to help improve the English ability of Japanese people, English language classes have been introduced from the fifth grade of elementary school (approximately 10 years old). The Ministry of Education, Culture, Sports, Science and Technology (MEXT) (n.d.a) stated that by doing so, it is aiming to foster “a positive attitude toward communication” among young students. In terms of classroom content, MEXT explains that instruction should help pupils “actively engage in communication in a foreign language.” Pupils should also “experience the joy of communication in the foreign language” and “actively listen to and speak in the foreign language.”

Within the secondary education system, for a long time students have taken English communication classes during both their junior and senior high school days (approximately 12-18 years old). In fact, MEXT states that at junior high school level, the aim is to “develop students’ basic communication abilities such as listening, speaking, reading and writing, deepening their understanding of language and culture and fostering a positive attitude toward communication through foreign languages” (Ministry of Education, Culture, Sports, Science and Technology, n.d.b). Many of these communicative classes are team-taught by a Japanese teacher in tandem with an assistant English teacher (a native English speaker from the Japan Exchange and Teaching (JET) program introduced in 1987). On paper, therefore, there is a strong commitment to communicative ability by the Japanese government.

However, as the high-stake national tests approach (which do not include a speaking component due to the large number of students), schools often replace communication classes with exam-related studies that focus on aspects of English such as grammar. This move away from communicative classes can be a demotivating factor for studying English (Kikuchi & Sakai, 2009, p.196). Why is there a move away from communicative classes at the senior high school level? It is because communicative ability is not part of the

national tests. When it really comes down to it, communication becomes low priority.

In spite of the structural problems that marginalize the importance of spoken English ability, English conversation classes do still exist in Japan, and often it is the non-Japanese teachers who are asked to teach these classes, whether they be in primary, secondary or higher education.

Oral tests for conversation classes

How should we test students in these conversation classes? There is a choice – a direct test (an oral test) or an indirect test (some form of non-oral test). Direct tests are when teachers test what they have actually been teaching and what the students have actually been learning. As Harmer (2007, pp.381-382) states, a direct test “asks candidates to perform the communicative skill which is being tested. Indirect test items, on the other hand, try to measure a student’s knowledge and ability by getting at what lies beneath their receptive and productive skills.” A presentation course would use presentations as the testing tool; an essay-writing course would make students write an essay. For a driver’s license, or a pilot’s license, it is important to test whether the examinee can actually drive a car or fly a plane. A simple paper test of theory would do little to make our roads or skies safe. Direct tests are logical therefore, if we want to test conversational ability, then, whenever possible, we need to make students take a direct oral test, not an indirect written or multiple-choice test.

Characteristics of conversation

Creating a concrete framework for communicative competence is difficult. McNamara (2000, pp.18-19) points out that measurements often include grammatical or formal competence (the knowledge of language), sociolinguistic competence (the use of appropriate language in different settings), strategic competence (the ability to compensate for “imperfect linguistic resources in a second

language”), and discourse competence (the ability to deal with extended language use in context). He concludes that, “attempts to apply a complex framework for modeling communicative language ability directly in test design have not always proved easy, mainly because of the complexity of the framework” (p.21).

In spite of the complex nature of communication, many of the components of conversations can only truly be tested by oral tests. Perhaps three of the most important components of conversation that act as reasons for using oral tests over other forms of tests when testing conversational ability are: that conversations take place in “real time”, conversation is face to face, and conversation is interactive and dynamic (Cornbleet & Carter, 2001).

(1) Conversation takes place in real time

Conversation is different from writing in that we face time pressure. If we receive an email question from a friend, we can start to answer, break off to do something else, return to our response, edit it, and finally check it carefully before we send it. This is quite clearly not possible in a conversation. If there is too much hesitation, then the listener begins to lose interest. This real time pressure of conversation can only really be tested in a conversation. The immediacy of the response is an important skill in conversation.

When we talk, we usually do it to “get our meaning over”. We don’t speak in the same way we write. We use incomplete sentences, simpler English, and we make grammar errors. It is fluency and communicative ability that are more important than grammatical accuracy. When a Thai student and a Japanese student speak English, the Thai student would not worry if the Japanese student omits a definite or indefinite article. Saying “I went to zoo” immediately, is more communicative than a grammatically perfect “I went to the zoo” that takes the speaker thirty seconds and several revisions to formulate.

Accuracy is, of course, important, but communicative competency is more important. Written tests that focus on accuracy

can actually deter students from trying to speak English, and teachers from trying to make them speak English. This is a widely observed phenomenon in English classes in Japan. The importance of “backwash” in testing should be recognized. Hughes, (1989, p.1) defines backwash as “the effect of testing on teaching and learning”.

If you test a conversation class with an oral test, there is more chance that the students (and teacher) will realize the relevance of classwork conversation exercises, thereby taking the speaking classes more seriously.

(2) Conversation takes place face to face

Apart from a few exceptions such as in telephone exchanges, conversation usually takes place face to face, and we can see many things from the facial expressions, gestures, and body language of the person we are talking to. We can gauge the person’s emotions, level of interest, and comprehension. We then react to these cues, by changing topic, allowing the other person to speak, or repeating something. These are important aspects of conversation that can only be tested with an oral test.

(3) Conversation is interactive and dynamic

Conversation is a collaborative act between two or more people. The social skill of turn-taking, knowing when to speak and when to listen, is important. We need to know when to use pragmatic devices to show our interest and willingness to participate in the conversation. Simple utterances such as “Really,” “Hmm,” or “Okay,” can show that we are involved. In conversation, we need to be able to respond to a dynamic situation.

All of these components are important in conversation, and it is important to test them with an oral test whenever possible.

A model for an oral test of conversation for university and college students

It is not that students in Japan do not know a lot of English, as they have studied it for 6 years, both at junior and senior high school. A glance at a typical university entrance exam shows the complex reading passages that students are expected to understand. The problem is that because speaking is not examined, it is also not practiced.

There are many ways of doing oral tests, each with advantages and disadvantages (see Underhill, 1987, pp. 22- 42, for examples), but the test outlined in this paper is for pairs of students, which also reflects how most textbooks practice conversation in typical English conversation courses at a Japanese college. The aim for the test is for two students, chosen minutes before at random, to converse for 3 minutes on a topic only revealed minutes before so as to avoid memorization, and to test proficiency at communication. This test can be supplemented with a “prepared spoken test”, such as a presentation, role-play or skit, which would reflect achievement and hard work, but this paper will focus on the “unprepared spoken test”, which it is hoped will encourage positive backwash and be a fairer test of conversational ability.

Step One: Classroom closed pair practice tests

The students should be informed at the beginning of the course that the final test will be a spoken one, and that it will be a conversation. The backwash effect is to make students realize that practice time in class is invaluable. Having a practice test in the third or fourth class, gives students a clear idea of what to expect.

Depending on students’ confidence and ability levels, it might be beneficial to have students do many closed pairs practices (without monitoring), before doing an ‘open class’ practice in front of the class. The teacher gives the students the first question such as, “What did you do yesterday?”, or “How was your summer holiday?” and then students have 3 minutes’ thinking time. This is for

formulating thoughts about the topic, and possibly to look up useful vocabulary items, but not for writing down sentences, which is what many students will want to do. Weaning students off memorization and reading prepared answers is essential in their development as spontaneous speakers of English.

Next, the students either find a partner or are allocated one, and then, using a timer, the teacher tells the students to begin. When the three minutes is up, the teacher stops the students, so that they develop a sense for how long the test will be. Pausing for a couple of minutes for students to reflect on what they said, and to attend to any problems they encountered, is a good way to foster autonomous thinking. Then, a repeat of the task, but with a different partner will help develop both confidence and fluency. Changing the partner means the questions asked will change, but a student's responses can be improved. (Changing the partner is also important as it prevents dependency on one person. Constantly working with a new partner means a student is exposed to a variety of voices and ability levels.) If students are very weak, it might be necessary to start with 60 or 90 seconds, and to gradually extend the time required.

Step Two: Open class practice test

For the first 'public test' practice, students are called at random and have a few minutes to ponder the opening question, which the teacher writes on the board. Before the test begins, it is important to insist on two rules. The first is that no Japanese (or other L1) is used. When conversing in a second language, miscommunication will be frequent, so solving any linguistic problem in English is a valuable skill to practice. The second rule is that students must keep talking for the time stipulated. This is critical as it places responsibility on the students. Although this might seem harsh, insisting that students continue speaking, by changing the topic if necessary, impresses on them that they need to take

responsibility for speaking for 3 minutes. If the teacher allows them to finish after 90 seconds then they will not strive to get better.

Developing autonomy

After the practice the students fill in a form (Figure 1) to help them improve both their own performance and to learn from their classmates.

<p>Things you did well.</p> <p>1</p> <p>2</p> <p>3</p> <p>Things which you can improve</p> <p>1</p> <p>2</p> <p>3</p> <p>The best speaker today wasbecause</p> <p>The most confident speaker wasbecause</p> <p>The most interesting conversation was betweenandbecause</p> <p>An expression/word I heard for the first time today was.....</p>

Figure 1: An example of a student evaluation form

Typically, Japanese students will be reluctant to write anything in the “Things you did well” section, so the teacher needs to insist upon it. Comments on the initial feedback form might include such things as, “I smiled”, “I could talk for 3 minutes”, and for some, “I enjoyed it”.

“Things which you can improve” might include basic aspects, such as, “speak more loudly”, “learn some more natural expressions”,

or “give longer answers”. In later feedback forms, it is hoped that the items mentioned in this section will be transferred to the first section as students attempt to improve their performance. These two sections provide a record for the student, that they can consult before practice tests, encouraging the targeting of an area in which they are weak.

The next questions focus students’ thoughts on what makes a good speaker by observing peers, and to think about what makes a good conversation too. These questions can be changed depending on what the class most needs. For example, for some classes it might be necessary to include, “Who had the easiest voice to hear?”, or “Which dialogue was the most natural? Why?” By highlighting the questions prior to the practice, students will work on areas that the teacher feels most need improvement. If the teacher is able, taking in these forms, writing some comments, making a photocopied record and returning the original to the students, shows each student that the teacher is taking a personal interest.

If, in most classes, the students have the opportunity to do a 3-minute private practice it is not necessary to overdo the public practice. However, occasional ones enable the teacher to see what is lacking in the class. Often it is manipulating the conversation that is lacking, and this can be rectified in subsequent class activities. For example, students will often talk as if it were a tennis match, with question and answer, followed by yet another question and answer. Practicing commenting on an answer (as shown in the example conversation below) is a useful activity.

Student A: What did you do in the spring holiday?

Student B : I went to Kyoto.

Student A: Oh, that’s a nice place. My grandmother lives there.

Interrupting, in which the goal is to stop one’s partner from finishing what they wish to say is an enjoyable activity, as is learning different ways of reacting with “Wow,” “Oh No,” “That’s fantastic,” “Is

that right?” amongst many that help make the conversation sound more natural. Showing one is listening also helps in that regard (including fillers such as “umm”, “err”) and practicing changing the topic when a particular subject has been exhausted. Students find it liberating to know that it is not necessary to complete each sentence, but to speak until the meaning has been understood. Before each 3-minute practice the teacher’s instruction that they need to concentrate on getting their point across rather than on grammatical accuracy is also liberating. As Brazil (1995, pp.26-27) commented, “...speech is characteristically used in pursuit of purpose...the practice of inventing a sentence...is a practice of the sentence grammarian, not the user”. Of course, accuracy is important, but it can be practiced in other activities. In the test, the goal has to be communicating, the student’s utterances need to be understandable, but the student should not be focusing on form but on meaning.

Step three: Mid-term Test

At the half-way point of the course a full practice test should be conducted, exactly as it will be done in the final exam. The students then will not be distracted by anything come test day. The only difference between the practice tests and the mid and final exam is that students evaluate their classmates. There are many ways of doing this, and depending on the level of the students it might be possible to define 3 or 4 criteria, such as “fluency”, “complexity”, “attitude” or “content”, but these can prove too difficult for many non-English majors, and result in unconsidered grading. For example, a weak student will not be able to grade a better student for complexity. Furthermore, with such an approach it is impossible to cover all aspects that make a successful communicator, and other features such as confidence, humor, manner, and voice quality, are just as important. An overall impressionistic grade works well for most students. To avoid students awarding too many high grades, rationing the number of A+, A, B+, or B, encourages students to

think more critically. If students write a comment next to the grade they award it is possible to see the criterion employed.

Figure 2 is an example of a class evaluation, in which students were permitted to award 4 A+, 4 A, 4 B+, and 4 B grades in a class of 26, and the others were left ungraded. The A+ grades have then been changed to a 5, A to 4, B+ grades to 3, B to 2 and those ungraded as 1. There is almost unanimity as to who the strongest and weakest speakers in the class are. This exercise also enables the teacher to see if the students have a differing view of any student and can make the teacher reconsider.

Student	Converted grades	
Student A	5 5	125
Student B	5 5 4 5 5 4 4 4 5 5 5 4 5 5 5 4 5 5 5 4 5 5 5 5 5 5	118
Student C	5 4 4 4 3 5 5 5 5 5 4 5 5 4 5 5 5 5 5 4 5 4 5 4	113
Student D	5 5 5 4 5 5 5 5 2 4 5 3 4 3 4 4 5 5 4 5 4 4 3 3 2	103
Student E	4 3 4 4 4 5 4 3 4 4 4 4 5 4 5 4 3 5 5 4 2 5 5 3 5	102
Student F	3 4 5 1 5 4 4 4 3 4 5 5 4 4 3 3 4 5 4 3 5 3 4 5 4	98
Student G	5 3 3 5 4 4 4 5 5 5 2 4 4 5 5 3 4 4 4 2 1 4 5 4 4	98
Student H	4 5 5 3 5 4 5 4 4 3 5 4 5 4 3 4 2 3 2 4 4 3 3 4 4	96
Student I	2 4 3 3 4 3 3 1 2 1 4 4 3 2 2 2 4 4 3 5 5 2 4 5 4	77
Student J	4 1 4 4 4 4 4 1 1 3 3 3 3 4 3 3 4 3 4 2 1 4 4 2 5	74
Student K	2 1 3 3 3 3 3 4 3 2 3 3 2 3 4 4 2 3 5 4 1 4 3 3 2	73
Student L	4 2 2 2 1 3 2 3 4 4 1 2 2 2 4 2 3 3 1 3 2 1 2 2 1	58
Student M	2 5 3 3 3 1 2 2 1 3 3 4 1 2 1 2 1 2 3 2 1 2 1 2 2	54
Student N	1 2 2 3 2 1 2 4 1 1 2 1 1 1 3 1 3 2 3 3 2 2 1 4 1	49
Student O	1 3 2 2 1 1 1 2 3 1 1 3 1 3 1 1 1 3 2 2 3 1 2 3 3	47
Student P	3 2 1 3 1 3 1 2 2 2 1 1 1 2 2 1 2 2 1 1 3 2 1 1 3	44
Student Q	2 2 1 1 1 2 1 2 3 2 3 1 2 1 1 2 1 1 1 1 3 3 1 2 2	42
Student R	3 1 1 2 1 1 1 2 1 1 2 2 2 3 3 1 2 1 1 3 1 1 3 1 1	41
Student S	1 3 2 1 1 1 2 3 1 2 1 3 1 2 1 1 3 1 2 2 1 1 1 1 2	40

Student T	1 4 1 1 2 1 1 1 3 1 1 1 1 1 1 2 1 1 1 3 1 1 1 3	36
Student U	1 1 1 1 2 2 2 1 1 1 2 2 3 1 1 3 1 2 1 1 2 3 1 1 1	36
Student V	1 3 1 1 1 3 1 3 2 2 1 1 1 3 1 1 1 1 1 2 1 1 2 1 1	36
Student W	1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 4 1 1	30
Student X	1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 4 1 1 1 1 1 1 1 1	29
Student Y	1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	27
Student Z	3 1	27

Figure 2: An example of class student peer evaluation

Class tests are not “high-stakes” tests (as opposed to entrance exams) so it is perhaps more important that the test is a learning experience for all students, than it is to get the most accurate grade. If teachers wished to be more precise, they could remove a percentage of the outliers for each student, but for most teachers’ purpose it is probably not necessary.

There are some disadvantages with peer grading. Falchikov and Magin (1997) mention possible gender bias, so that in a class which is predominantly female, males might be under-graded and vice-versa. Abson (1994) points out that it can lead to discrimination against certain individuals, so teachers need to be alert to this. With less mature students it might happen that groups of friends decide to inflate each other’s scores, but this is easy to spot and exclude. In Japan, with the “sempai-kohai” (senior-junior) relationship so important, students who repeat a year might be over-graded by virtue of seniority, so that too, needs monitoring. If there are obvious problems in the mid-term test the teacher has the opportunity to talk to students and address the difficulties.

Steps Four and Five: Further closed pairs and open pair tests

Between the mid-term test and the final test, students continue with further closed pair practice in which they focus on improving any weaknesses highlighted in previous feedback forms. It might be useful to hold one more open-class practice test, and for students to fill in the appraisal form. By the latter stage of the course students should be making more nuanced comments on their own performance and those of their peers.

Step Six: Final test

This is a repetition of the mid-term test. Students are notified of the opening question for the dialogue, and have 2-3 minutes to think, but not write, about what they could say. Whilst they are thinking, each student draws a number (from a hat is enjoyable) at random. The students might then perform as numbers one and two together, or the first and the last numbers together; it is not important as long as it is possible for all to see the pairs are not fixed, and that there is no favoritism. The teacher might remind the students that those who go first are at a slight disadvantage as they have less planning time and more stress. It is also a good idea for students to initially record impressions rather than final grades, and to make the necessary adjustments once the last pair has finished. Students should not talk until the grade sheet has been collected. For an uneven number of students in the class, it might be necessary for one student to perform for a second time. The teacher could ask for a volunteer or nominate someone who either did unusually poorly, or someone who the teacher feels is on the borderline between two grades.

The students' grades would account for 50% of the final grade, with the other 50% being awarded by the teacher.

Advantages of this test

It is a direct test that tests the students' conversational skills. The students know that they will have to speak in the test, so the backwash effect is that students should use practice time to develop their speaking skills. The test develops more student autonomy, as they adopt a more critical approach to both their own and classmates' performances. Noticing how language is used, and being encouraged to incorporate language used by others are also benefits of this approach. The best students develop the ability to think on their feet, which makes them confident that they could manage a conversation in the real world should it be necessary. Those who have taken the opportunity to practice regularly make great strides with fluency and can turn thoughts into words. Students learn how to manipulate a conversation, to change the topic if it is dull, and to interject when there is something to say. Less important linguistically, but impressive nevertheless, is how attentively students listen to each other, and that it can help bond the class to hear peers talk about their lives and also, as is sometimes the case, to laugh together. Another positive aspect is that it is not a test that can be revised for in a traditional sense. There is nothing to memorize or study. The only way to do well is to practice speaking on any topic, which is the only way to become a fluent speaker.

Disadvantages of this test

In the test outlined here only one very open-ended question is used in the test. Typically, the test day question will be something like, "How was your year?" or "What's new?" These are deliberately broad topics so that students can talk about something which is of interest to each individual. Initially, students might copy the previous pair in content, but by stressing the importance of originality in dialogue, students can avoid this trap. It is possible to have a few different topics and switch between them if a teacher feels the test would be fairer with a variety of topics, but it is important that the questions are of similar difficulty. Some topics can lead to low level

exchanges, such as “What are your hobbies?” in which there can be endless, “Do you like ...?” questions, so the topic must be chosen with care. Avoiding topics done in class is important, unless the teacher wants the test to be one of achievement rather than proficiency.

There is an element of luck in determining both a student’s partner, and the timing of when students perform. It is possible for a weak student to limit the degree of difficulty a higher-level student could use when talking, but conversely, because the weaker student realizes this, it can lead to a greater effort. If the pairs were not drawn at random, but by ability instead, this would be beneficial for the better students, but it might stigmatize the weaker students who are paired together. Students who do not speak till later in the test have had the opportunity to listen to others answer the same question, so can use ideas and, hopefully, language used by other students. In a high stakes test this would not be fair, but in the classroom setting, it fosters the behaviors of autonomy that many teachers would like to encourage.

Having a conversation with 20 other people listening is obviously a false environment, but all speaking tests are contrived to some extent. For example, a test of conversation with just a teacher and one student has the problem of power-distance, in which the student finds it very hard to question the teacher as an equal (Kormos, 1999). A pair of students and just a teacher might work better for shier students but the class would be denied the learning opportunity of seeing others perform and the chance to grade. There is also a certain enjoyment to be had from performing (well) in front of classmates.

Conclusion

If, as teachers, we want our students to be able to communicate with others in spoken English, they must have the opportunity to practice this skill. There are unique pressures in conversing, as Brazil (1995, p.11) reported. “As speakers, we know

that causing it (speaking) to happen is not always without its problems: our ability to put together what we want to say may not always be equal to the pressure to keep up with ourselves, so to speak, in the delivery of our message. As listeners, too, we frequently feel ourselves under similar pressure. The fact that time is passing makes it imperative to decode what we hear promptly so as not to miss what comes next.” The conversation test outlined in this paper attempts to replicate the pressures Brazil observes, and by meeting the challenge successfully (which nearly all students can manage after concerted practice), the fear of conversing gradually recedes. Bearing this in mind, whenever it is practically possible, oral testing is highly recommended for conversation classes. Furthermore, using a framework such as the one outlined in this paper that incorporates self and peer evaluation, enhances the development of student autonomy.

The Authors

David Evans is an assistant professor of English at the National College of Nursing, Japan. His research interests include oral testing, self and peer evaluation, and extensive reading.

John Herbert is a professor in the Faculty of Liberal Arts, Hiroshima Jogakuin University, Japan. His academic interests lie in the fields of TEFL and sociolinguistics.

References

- Abson, D. (1994). The effects of peer evaluation on the behavior of undergraduate students working in tutorless groups. In H.C. Foot, C.J. Howe, A. Anderson, A.K. Tolmie & D.A. Warden. (Eds.). *Group and Interactive learning*. UK: Computational Mechanics Ltd, pp. 153-158.
- Brazil, D. (1995). *A Grammar of Speech*. Oxford: Oxford: Oxford University Press.

- Cornbleet, S. & Carter, R. (2001). *The language of speech and writing*. London: Routledge.
- Crystal, D. (2007). *How language works*. London: Penguin.
- Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group process work. *Assessment and Evaluation in Higher Education*, 22 (4), pp. 385-396.
- Gottlieb, N. (2005). *Language and society in Japan*. Cambridge: Cambridge University Press.
- Harmer, J. (2007). *The practice of English language teaching*. Harlow: Pearson Longman.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kormos, J. (1999). Stimulating conversations in oral-proficiency assessment: A conversation analysis of role plays and non-scripted interviews in language exams. *Language Testing*, 16 (2), pp. 163-188.
- Kikuchi, K., & Sakai, H. (2009). Japanese learners' demotivation to study English: A survey study. *JALT Journal*, 31 (2), pp. 183-204.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford Univ. Press.
- McVeigh, B. (2002). *Japanese Higher Education As Myth*. New York: M.E. Sharp. Ministry of Education, Culture, Sports, Science and Technology (MEXT). (n.d.a). Foreign language activities. Retrieved February 6, 2013 from http://www.mext.go.jp/component/english/_icsFiles/afieldfile/2011/03/17/1303755_011.pdf
- Ministry of Education, Culture, Sports, Science and Technology (MEXT). (n.d.b). Foreign languages. Retrieved February 6, 2013 from http://www.mext.go.jp/component/english/_icsFiles/afieldfile/2011/03/17/1303755_013.pdf
- Underhill, N. (1987). *Testing Spoken Language*. Cambridge: Cambridge University Press.