

# Oral Testing in the Communicative Classroom

Carolyn Saylor-Lööf

*Kansai University*

Rebecca Calman

*Kyoto University of Foreign  
Studies*

## Reference Data:

Saylor- Lööf, C., & Calman, R.. (2005). Oral Testing in the Communicative Classroom. In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.) *JALT2004 Conference Proceedings*. Tokyo: JALT.

This article looks at the oral testing of conversation skills in the context of communication classes for Japanese university students. An overview of oral testing is given, and it is argued that testing students in conversational pairs with one another is better than an interview test in which students are paired with an examiner. Furthermore, the many benefits of giving an oral test are discussed, and a model for preparing and administering such a test is proposed.

本稿は大学のコミュニケーション・クラスでの実施されているオーラル・テストを取り上げている。オーラル・テストの概要を述べた後、先生が学生をインタビューする形式のテストより学生同士がペアになって会話をするという形式のテストの方が優れていることを論じる。さらに、オーラル・テストの有益性について述べた後、オーラル・テストの準備、実施方法を提案する。

**J**apanese university students have all had six years of English instruction in secondary school and, thus, possess a basic understanding of the language. On the whole, their written English is better than their spoken English; they have had little experience with English as a living language. Therefore, the primary goals of an English communication course are to teach English as a linguistic-cultural-social unit, to facilitate learners' use of the language as a tool for communication, and to "actively develop students' ability to communicate in a socially appropriate manner" (Kurzweil, et al., 2002, p. 32).

To realize these objectives in the classroom, we have developed a syllabus designed around conversation skills, such as: opening and closing a conversation; asking follow-up questions; giving long answers (i.e., details and more information); opening and changing topics; using active listening, which includes rejoinders (e.g. "I see," "Really?") and backchannelling (e.g. "Uh-huh," "Yeah"); turn-taking; eye contact; body language and key prosodic features. These skills are taught together with other necessary components of language so that students can practice speaking about topics relevant to them: personal information, family, hobbies and interests, school life, friends, work, future plans, past events, and so on.

At the beginning of the course, students are told that they will have a conversation test at the end of the semester and that everything done in the course directly pertains to acquiring the skills necessary to perform successfully on this test.

The oral test, then, not only is a natural complement to the semester's coursework, but plays an important role throughout the course itself. We would, therefore, like to not only present our testing procedure but examine its effect on and implications for our learners and our course as a whole. To do this, we find it useful to begin with a review of the theory behind and current issues in the field of oral testing.

### Theories and Issues in Oral Testing

Oral tests are used to test speaking ability and traditionally have focused largely, and often solely, on linguistic proficiency (Hughes, 2003). Recently, however, there has been an increase in the inclusion of conversational skills and strategies, as well as para- and extra-linguistic features of the target language.

Oral tests serve many purposes: to measure language proficiency; to assess achievement of the objectives of a course of study; to diagnose learners' strengths and weaknesses; to identify what they do and don't know; and to assist in the placement of learners within a teaching program (Hughes, 2003).

The most common format of oral testing is the interview, in which the test taker converses with an interviewer and his or her performance is evaluated. There is often an assessor present who does not take part in the spoken interaction but listens, watches, and evaluates the abilities of the test taker. If there is not a second person available to act as an assessor, then the interviewer must also assess the candidate's abilities.

Although the oral interview is widely practiced, there has been mounting criticism against its use in recent years. Many researchers agree (e.g. Bachman, 1990; van Lier, 1989; Lazaraton, 1992; Young, 1995) that the oral exchange that occurs between an interviewer or tester and test taker does not mirror or even closely replicate natural, or real-life, conversation. As Johnson and Tyler (1998) observe, in natural conversations topics as well as turn distribution, order and length are mutually negotiated by both interlocutors; however, in an interview test, they are primarily "set in advance and controlled by the testers" (p.48). They further argue that "naturally occurring conversation is by its very nature interactive, and that a crucial part of this interactiveness is a sense of involvement or reactivity among interlocutors. [They] have noted that...testers' contributions consistently lack this quality of conversational involvement" (1998, p.48). Indeed, a real-life conversation is a spontaneous creation between two (or more) involved participants which has not been planned ahead, and the content, sequence, and outcome are largely unpredictable.

Educators and testers are rightly concerned that speaking tests should mirror real life speaking situations. As Bachman states, we need to "capture or recreate in language tests the essence of language use, to make our language tests 'authentic'" (1990, p.300). He proposes a 'real-life' approach which aims to develop authentic tests and which is primarily concerned with: face validity, how the test appears to and affects those taking it; and predictive validity, how accurate the test is in predicting future non-test language ability.

Another concern with interview tests, as Kormos (1999) points out, is that they are often unequal social encounters and, therefore, “inherently resemble interviews rather than natural conversation” (p.164). She further raises the issue that “the schemata for participating in interviews might be culturally different” (p.171). Young supports this proposition by finding that Asian test takers will expect the interviewer to lead and dominate the conversation, thus obscuring the test-taker’s discourse abilities (1995).

One can conclude that in an oral test in which the test taker’s partner is a trained interviewer, the social inequality between the two interlocutors, the lack of spontaneity and involvement on the part of the tester, and the fact that the conversation held is not ‘authentic’, will hinder the performance and assessment of conversational competence.

### **An Alternative to the Oral Interview Test**

In recent years there has been considerable criticism of the oral proficiency interview. Although a review of the literature has found many shortcomings to the interview format, there do not seem to be many solutions or alternatives to this format proposed. We would like to offer our oral test as one alternative to the interview format.

The distinguishing feature of our test is that it is not an interview. Students are examined in pairs, so that each student has another student for a partner during the test; the teacher is only an assessor. There are many advantages to using this arrangement. First, the test takers are socially equal, which not only ensures a degree of comfort, but allows them to express themselves more freely. We have

found that this is especially relevant for Japanese students, as conversational style varies greatly depending upon the relative social status of the speakers. If a student speaks with someone older, for example, the student will not participate equally in the exchange but will instead assume a lesser role, allowing the elder to lead and dominate the interaction. Furthermore, a student-student interaction is much closer to a real-life conversation. A further advantage to student pairs is that it is easy for them to find common topics to talk about. For these reasons, we believe our oral test to have a high degree of authenticity, in that it closely resembles a real-life conversation.

It is always a challenge to design a test so that a natural conversation can occur. Learners must be relaxed and confident enough so that a conversation can spark, topics flow and, thus, “allow the activity (the conversation) to become dominant, and its ulterior purpose (a language test) to be temporarily subordinated. The oral test then reaches its highest degree of authenticity by no longer being a test” (Underhill, 1987, p.45). We propose that in our oral test, we provide the necessary conditions to create such a situation conducive for having natural conversations.

Other factors that contribute to the authenticity of our test are preparedness and resulting confidence. Throughout the semester, students are consistently working on aspects of conversational competence, leading up to the final oral exam. As our test mirrors our course content, students are likely to be well prepared and thus have confidence when taking the test. Our students are also well informed by their teacher as to what the actual test conditions and procedure will be like. The more information and preparation they have prior to the

exam, the better they will perform and the more accurate a picture we will get of their oral abilities.

Our speaking test not only serves as a way to measure our students' conversation abilities (i.e., achievement in the course), it has many other purposes as well. The main reasons for having a final oral test are that it motivates and focuses the students throughout the course, provides a framework for both teaching and learning, and gives students a clear direction and goal to work towards. It encourages them to pay better attention in class, study harder, and in general, take their learning endeavor more seriously. This effect of the test on teaching and learning is called washback by some (Underhill, 1987; Bachman, 1990; McNamara, 2000) and backwash by others (mainly, Hughes, 2003). We find the washback of our test to be extremely beneficial and useful.

Another advantage of our oral exam is that the test in itself is often a positive learning experience -- even enjoyable -- to many of our students. It gives them a sense of accomplishment. After the test, students often express surprise at how well they performed during the test, or that they enjoyed the conversation they had with their partner, or that they forgot they were even taking a test. Of course, we also receive feedback where students express regret that they did not perform as well as they had expected, or that they did not prepare enough, as well as students' realizations of their weaknesses and vows to work harder in the future. Another purpose, then, of our test is diagnostic: it tells the learners as well as the teachers what the students' strengths and weaknesses are and what areas and skills need to be worked on more in the future.

To attain a clearer understanding of how the test is actually carried out, we take a further look into the details of the examination procedure.

## Planning and Administering a Peer Conversation Test

How should a teacher organize the oral testing of conversation skills? The authors were interested in creating a usable testing procedure that would allow a teacher to examine 30 students during one 90minute class period with reasonable reliability. The testing process they propose includes preparation, both psychological and pedagogical, a marking system, testing procedures and criteria, and feedback or follow up to the oral test.

### Preparation

In the first week of class, the teacher announces that there will be an oral test of conversation skills at the end of the semester. Students are reminded of the test and may be told to take note of material for the test periodically throughout the course.

A week or two prior to the actual event, students receive a written guideline to study for the conversation test (see Appendix I). This can be done on the board as well as on paper. Also, a week prior to the actual test, students take a mock conversation test (see Appendix II) in which they observe and evaluate one another. A mock test gives students insight into the evaluation process, as well as an opportunity to review the semester's material, practice the specific skills being tested, and lower anxiety about the upcoming test.

### *Test Criteria*

The criteria for the oral test are basic conversation skills outlined at the beginning of the course and taught during the term. These include starting a conversation, giving long or detailed answers, asking follow-up questions, using rejoinders and other active listening skills, opening and changing topics, and closing the conversation. These core conversation skills are the primary consideration in evaluating students' oral tests. Things like turn-taking, the use of small talk, using repair skills to clarify or ascertain meaning, making smooth transitions between partners, employing appropriate language, and using culturally acceptable conversation topics might also be included if they have been taught in class.

While not primary to the exam, other factors that may be considered include body language, specific vocabulary, personal interaction skills, prosody, and clarity of expression. Appropriate body language, such as eye contact, facing a conversation partner, nodding, smiling, and gesturing may be considered when evaluating student performance on the conversation test. Of course, students were expected to speak exclusively in English, and to the best of their ability, use natural intonation and pronunciation.

The oral test is criterion-referenced. That is, our oral test shows whether the students are able to perform conversational tasks satisfactorily. This is in contrast to norm-referenced tests in which students' abilities are judged in comparison with one another.

### *Test Administration Options*

How do teachers organize their oral tests? Testing 30 students in a 90 minute class period apportions three minutes per student, without allowing time for activities such as moving or taking attendance. Also, while the teacher observes oral tests, the other students need to be occupied or out of the way. In our oral test model, students are tested apart from the class, in order to lower anxiety and increase concentration. This is supported by Underhill who contends that learners can relax in ordinary surroundings, such as a hallway or a cafeteria, which de-emphasize the test-taking aspect of the conversation. This helps the students speak more naturally (1987).

The system we recommend, and have used successfully, is to test the students in double pairs. That is, the teacher simultaneously assesses two pairs of students, each pair conducting a separate conversation. The teacher sits in the middle of the two pairs. Students are tested in ten-minute intervals, allowing about 5 minutes for the actual test, and five minutes for coming and leaving, getting settled, brief feedback and the teacher's final scoring. In this fashion, 30 students can be observed in the 90 minute class period. In addition to the timing advantage of the double-pair system, having two pairs speak at once greatly reduces students' stress because a) they feel they are not the sole object of the teacher's focus, and b) it more closely mirrors a classroom atmosphere wherein other students are talking at the same time, which they are used to.

What is the rest of the class doing while pairs of students are being tested? One possibility is the teacher sets the class a task and takes four students at a time to a nearby empty

classroom or the hallway, where the oral test is administered. Students in the class can be assigned to write an essay reflecting on the semester or some other task that will engage them sufficiently to keep them occupied and quiet, and one in which cheating would not be an issue, as they must work unsupervised.

Another possibility is that students come by appointment. Several students should be scheduled to arrive at a specific time, for example eight students come every twenty minutes. Once they arrive, they are randomly put in pairs, for example, by handing out cards. Four students then enter the classroom to be tested, and four wait their turn quietly in the hallway for 10 minutes. Having eight students scheduled for a specific time ensures that they cannot pre-plan a conversation with one partner before the test. They can, though, study in a large group together and become familiar with one another which are beneficial for lowering anxiety during the actual test.

### **Test Task Options**

Depending on what topics and skills have been worked on throughout the course, the teacher will need to set a test task accordingly. The two pairs of students being simultaneously tested should be given different test tasks so that they are not encouraged to listen to the other pair's conversation and they can better tune it out. The task that we set in the first semester test is as follows: one pair is told that they do know each other, and one pair is told that they don't know each other and are just meeting for the first time. Their resulting conversations should be quite different. In the second semester test, each pair selects one of several face-down

topic cards, reflecting the topics covered during that semester (e.g. traveling, health and illness, food, cultural differences, family and friends).

### **Evaluation**

The teacher must have an efficient evaluation system, particularly because two pairs are being evaluated at once. The evaluation sheet the teacher uses will of course depend on the skills being tested and on how the teacher feels most comfortable organizing their evaluation system. We believe that it should closely resemble the mock test evaluation sheet which the students have used and are familiar with (see Appendix III for an example sheet). On this sheet, teachers can write evaluative numbers in the boxes, or just simply tick off skills. For example, every time a student asks a follow-up question, the corresponding box gets a tick.

In the event that one student fails the test or performs so poorly that it affects his/her partner, the student can be retested with another partner.

### **Feedback**

Students should get some immediate feedback after the test, ranging from encouraging comments like "good job" to handing the student an evaluation page with a final grade. General feedback to the class as a whole, as well as going over common mistakes and items for further practice, can be given in a class meeting following the oral test.



## Conclusion

The teachers who use the testing method we have described, as well as the students who take the test, have all given us very positive feedback.<sup>2</sup> Does this in itself mean, though, that our test is valid and reliable?

First of all, it is important to keep in mind that “oral tests are not like lists of questions on paper; they do not exist separately from the people who take part in them,” and it is therefore difficult to evaluate test validity and reliability (Underhill, 1987, p.107). There is always going to be some degree of subjectivity in oral testing.

The oral test method we propose is to some extent objective. Students are required to perform conversational tasks in order to pass. A teacher can reliably record that a given student has asked, for example, follow-up questions. However, the quality of the follow-up questions requires the teacher’s judgments, and is, therefore, subjective. Underhill calls this impression-based marking. This very subjectivity may benefit the testing process: “Deliberate and careful impression-based marking is the most direct and authentic reflection of this real-life process that is possible to have in an oral test” (1987, p.101).

In the simplest possible terms, test validity poses the question, does the test measure what it is supposed to? Our oral exam method can answer “yes” to this question. Students practice conversation skills in English and then produce these skills during a test. The oral test is designed to check that they have mastered a specific group of skills which are observable. Thus, the test has content validity.

In addition to the content being valid, we believe our test also has a high degree of predictive validity, although this is an area for further research. As the conversations that occur during the test closely resemble real-life interactions, the test should rather accurately predict how successful the students will be at using the language in similar future situations. We further conclude that the test has high face validity because feedbacks from teachers using this method and comments from students have been very positive.

There are many areas for future investigation and research which would help to evaluate our test’s reliability. First of all, teachers could compile and compare all their criteria schemes and grading systems, and look at actual marks given on individual criteria and overall grades, in order to see how consistent markers are with themselves and compared to others. Moreover, this process would enable us to see which criteria are the easiest and which are the most difficult to mark consistently.

Other ways to evaluate test reliability would be for teachers to observe one another’s testing process or to use one another’s criteria and marking systems while testing. In addition, teachers could test one another’s students or pair students from different teachers’ classes for a conversation test. As our test reflects the course as a whole, ideally, students should be able to perform satisfactorily no matter which teacher evaluates them or who their partner is. However, this would also involve human variables, as the students would be asked to perform in front of a teacher strange to them, and with a partner that they were not acquainted with; both factors could influence test performance.

Another method for determining reliability would be to videotape students having a conversation. This video could be scored by other teachers, and scores compared to check marking consistency.

Such investigations, however, lie beyond the scope of this paper. In the future, teachers may be inspired to experiment with various styles of test administration. Hopefully, our method can provide ideas and choices for testing students on an oral conversation syllabus.

### Notes

- 1) This article has been adapted from the following: Saylor-Lööf, C. & Calman, R. (2004). Oral Testing for Conversation Skills. *Kansai University Forum for Foreign Language Education* (3), 1-16.
- 2) All 12 teachers in the Communication Program at Kansai University used this method, by choice, and were very positive about it. In an anonymous end of term evaluation given by one teacher to her 300 students, 92% of students answered “yes” to the question, “Do you think the final oral test was fair?” In response to the question, “Do you think that what you learned in class prepared you for the final oral test?” 89% of students responding answered “yes.” And to the question, “Which do you like better, to be tested two pairs at a time or one pair at a time?”, 100% of students answered that they prefer the double-pair testing.

### References

- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Calman, R. & Walker, K. (2003). Learning Journals: Various Approaches Used in a University Level EFL Setting. *Kansai University Forum for Foreign Language Education*, 75-91.
- Doodigian, J. & Famularo, R. (2000). The English Communication I Program; Curriculum Development and Implementation. *Kansai University Kenkyuu Sentaa-hoo*, 215-235.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Johnson, M. & Tyler, A. (1998). Re-analyzing the OPI: How Much Does It Look like Natural Conversation? In R. Young & A. Weiyun He (Eds.), *Talking and Testing*. Philadelphia, PA: John Benjamins B. V.
- Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: a conversation analysis of role plays and non-scripted interviews in language exams. *Language Testing*, 16, 163-188.
- Kurzweil, J. et al. (2002). Communication 1 Syllabus: Designed by Consensus. *Kansai University Forum for Foreign Language Education*, 31-43.
- Lazaraton, A. (1992). The structural organization of a language interview: a conversation analytic perspective. *System* 20, 373-86.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.



- Tarone, E. (1998). Research on interlanguage variation: implications for language testing. In Bachman, L. and Cohen, A., (eds.), *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.
- Underhill, N. (1987). *Testing Spoken Language: A Handbook of Oral Testing Techniques*. Cambridge: Cambridge University Press
- van Lier, L. (1989). Reeling, writhing, drawling, stretching and fainting in coils; oral proficiency interviews as conversation. *TESOL Quarterly* 23, 489-508.
- Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning* 45, 3-42.

## Appendix I. Example of study sheet

### Useful Expressions for the Conversation Test

Greetings and Introductions	Hi! I'm (your name). Nice to meet you. Nice to meet you too. [Use these two only if you do not know your conversation partner.]
	Hi! How are you? Good morning! Hey! How's it going? Good afternoon! Hello! How have you been? Good evening!
Starting the Conversation	(With a person you do not know) (With any person) So, where are you from? So, how do you like this So, what's your major? English class? (With a person you do know) So, what do you think about So, how was your weekend? the Hanshin Tigers this So, what are your plans for this season? summer?
Passing Back the Question (Keeping the Conversation Going, Part One)	How about you? What do you think? What about you? What's yours? (In response to And you? "What's your favorite . . . ?")
Rejoinders (Keeping the Conversation Going, Part Two)	Surprised      Happy      Sad      Interested No way!      That's great!      That's too bad. Oh, yeah? You're kidding! Wow! Oh, no. I see. Really?! Wonderful! Uh-huh. Wow! I am sorry to hear that.
Follow Up Questions (Keeping the Conversation Going, Part Three)	How often do you . . . ? How do you . . . ? When did you start . . . ? Where do you . . . ? What do you like about . . . ? Who do you . . .with ? Why do you . . . ?
Long Answers (Keeping the Conversation Going, Part Four)	No, I haven't. But, I would like to. . . Yes, I did when I was . . . No, I don't. But, I do have . . . Yes, I do. I have had . . . No, I am not. I am . . . Yes, I am. I have been . . .
Finishing the Conversation	Well, I have to go now. It was nice talking to you. Well, it was good to meet you. I'd better go. O.K., Take care! You too! See you later!

You may NOT look at this paper during the Conversation Test.  
Please refer to class notes and worksheets for many more useful expressions.

## Appendix II. Mock Test for Students

### Conversation Evaluation Sheet

Listen carefully to a pair’s conversation. Evaluate their conversation abilities by filling in the boxes below with a number and with COMMENTS. Try your best to help your fellow classmates become aware of and improve their speaking skills.

Use these numbers to evaluate conversation skills:			
5 = very good / excellent	4 = good	3 = OK but could be better	
2 = needs improvement		1 = not good	
Pair 1		Pair 2	

	Name:	Name:	Name:	Name:
Greeting/ Introduction				
Start the conversation., (“So...”)				
Keep it going: Follow-up Qs				
Long Answers & Opening Topics				
Active Listening				
Turn-taking: Balanced Conversation.				
Finish the Conversation.				
Voice Volume				
Intonation				
Pronunciation (Katakana English)				
Eye Contact				
Appropriate Body Language & Gestures				
Content: Knowledge of topic	(N.B.: This category is only applicable for test on topics)			

Additional comments, suggestions for improvement:

Appendix III.

Evaluation Sheet for Teachers

Pair 1

Pair 2

Test Task:	(for example): DO know each other		DON'T know each other	
	Name:	Name:	Name:	Name:
Greeting/ Introduction				
Start the conversation. (“So...”)				
Keep it going: Follow-up Qs				
Long Answers & Opening Topics				
Active Listening				
Turn-taking: Balanced Conversation.				
Finish the Conversation.				
Voice Volume				
Intonation				
Pronunciation (Katakana English)				
Eye Contact				
Appropriate Body Language & Gestures				
Content: Knowledge of topic	(N.B.: This category is only applicable for test on topics)			