

An Analysis of Oral Testing in Pairs

Harumi Yamada

As English language teaching in Japanese educational settings shifted its focus more towards fostering students' communicative abilities, the need to directly assess their spoken English abilities through oral tests has increased in recent years. If we are to use oral tests as achievement tests, we need to know what actually happens when students converse during oral tests. Although quite detailed qualitative analyses have been carried out on such proficiency tests like those of UCLES (Lazaraton 2002 and others), little research has been conducted on the practice of oral tests in Japanese schools.

This study will show what actually happened during an oral test given at the end of a pre-intermediate general English university course. Quantitative and qualitative analyses of the language produced during the test will be carried out. Quantitative analyses will be carried out as to the number of turns, gender difference, mean length of utterance (MLU), test marks given, and partner familiarities. In the qualitative analysis, different features of discourse will be highlighted. This will provide an objective look into what could happen in oral tests carried out as part of achievement testing at schools.

Key words: oral testing in pairs, achievement test, empirical research into the nature of the discourse

1. Introduction:

The first time the word 'communication' appeared in Japanese Ministry of Education's the "Course of Study¹" was in 1989. As the Japanese business sector began to realise the necessity of increasing the number of Japanese who can communicate using English as an international language, the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) shifted its focus in the revised "Course of Study" in 2003 more onto cultivating students' practical communication abilities. In the overall objectives for foreign languages in the 2003 Course of Study for both junior and senior high schools, the key words are 'to develop students' practical communication abilities' and 'to foster a positive attitude toward communication.' At the university level, the MEXT University Council Report (1998) stated that 'it is important to develop human resources who can play active parts on the international stage. Thus, it is necessary to enhance foreign language education, ... and help students cultivate an ability to clearly express oneself through discussions, oral presentation, oral reporting and the like.'

In this situation, the need to directly assess students' spoken English abilities through oral tests has increased at schools. Since the introduction of Oral Communication subjects at senior high schools in 1992², more schools have started to practice oral tests to assess the students' abilities.³ At university level, more and more universities are practicing oral testing in their general English courses.

However, there has been lack of teacher training and study in this area (Yamada, 2004). Just as Lazaraton (1992) states that 'until recently few attempts have been made to look "inside" the language assessment interview in order to examine just what kind of speech event it is, an endeavor that would allow us to identify and describe performance features that determine the quality of conversational interaction (p373),' few attempts have been made to look 'inside' the oral testing at schools. In other words, there has not been much research to 'identify and describe performance features' (van Lier 1989:497) of the discourse the students display in an oral test. This study is an attempt to conduct empirical research into the nature of the discourse presented in a paired oral test carried out to assess the students' achievement levels. The writer hopes that this study will help to pursue the possibility of successful implementation of paired oral tests as part of achievement tests at Japanese schools.

This study examines paired oral tests as the object of analysis for three reasons. Firstly, Yamada (2004) found that 95% of her respondents (senior high school teachers in Fukui Prefecture) always or often use pair-work and group-work. Therefore, if the achievement test takes the paired or group format, then there would be consistency between the classroom activities and the method of assessment and productive washback could be expected.⁴ Secondly, the paired format gives students more chances and time to talk than 'an oral presentation' or 'an oral interview.' If we find this format to be valid and reliable, it would be more practical in a school setting where teachers have to assess many students at a time.⁵ Thirdly, by using this format, we can assess the students' competence to carry out interactive communication directly.⁶

2. Literature review:

Not so much qualitative and quantitative research has been done into school practices of oral testing itself nor into oral testing in pairs. This section will review the research on the paired format in oral proficiency tests and discuss the advantages and disadvantages of oral testing in pairs.

2.1 Research on the paired format in oral proficiency tests:

One of the most controversial points in the paired format in oral testing is the differences between candidates' language proficiency. Ildiko's (2001) quantitative study showed that differences in partners' proficiency levels had no significant effect. Nakatsuhara (2005) also reports that 'the presence of different proficiency-levels may not be as serious a concern as anticipated.' On the other hand, Norton (2005, 2006) and Lazaraton (2006) discuss how

pairing of candidates may have impact upon the language sample produced.⁷

There are some studies that indicate the appropriateness of the paired format for assessing the candidates' communicative competence. In a quantitative comparison of the paired and one-to-one formats Taylor (2000) found 'a more balanced interaction' and 'a richer and more varied sample of spoken language' in a paired format than in one-to-one interview tests. The qualitative comparison showed that the paired format was capable of eliciting more communicative language functions (Taylor *ibid*, and Taylor 2001). Galaczi's (2003) study indicated 'that the interaction which is the most conversation-like from a discourse perspective (collaborative) was also graded the highest by the FCE examiners (p.22).'

Lazaraton (2002: 152-160) summarizes her research into candidates' behaviour and the speech functions employed by them. She points to the need of devising test-tasks so that the speech functions we intend to elicit can actually be elicited in the tests.

2.2 Research on advantages and disadvantages of oral testing in pairs:

Advantages and disadvantages of oral testing in pairs have been widely discussed (Gyorgyi Egyud & Philip Glover : 2001, Foot : 1999, Norton : 2005, Csepes Ildiko : 2001, etc.). They are summarised as follows:

Advantages	Disadvantages
<ul style="list-style-type: none"> (1) Levels of stress are reduced because there is little power relationship in this format. (2) Candidates are able to give and receive help by cooperating with their partners. (3) Pairings help to produce better English than the one-to-one format. (4) A wider range of speech acts can be seen in this format. (5) This format supports good teaching. In other words, it has a good washback effect. (6) Grading is easier for the examiner. (7) It is non-threatening as the examiner is not seated directly opposite a student. 	<ul style="list-style-type: none"> (1) There is a lack of evidence in support of pairings. (2) Some students don't like pairings. (3) There are doubts as to whether the paired format is appropriate at lower levels. (4) It is possible that inequality of partners in a paired format could cause problems. (5) Unsuccessful pairing could result in the failure of the assessment. (6) There are doubts about scoring reliability.

These discussions are based on several research findings, but not much qualitative and quantitative research into school practices has been done to date.

3. Study:

3.1 Research questions:

In this study, we are going to try and answer the following two questions:

- (1) What could a paired oral test as an achievement test display?
- (2) What can be assessed by using paired oral tests as achievement tests?

3.2 Data collection:

3.2.1 About the subjects:

The subjects here are 22 university students of pre-intermediate level of English, who have taken 14 weeks of instruction (90 minutes a week) on a general English course aimed at developing communication abilities in all the four skill areas of speaking, listening, reading and writing. In class, the students did extensive pair work.

3.2.2 Regarding the data:

Three kinds of data have been collected here. One is the transcription of the video-recorded oral tests the subjects took. The second was taken in the form of a questionnaire, which was given immediately after the test in regard to partner familiarity and impression of test difficulty. In the questionnaire, the subjects were also asked to write down anything they thought about the test. The third is retrospective data. The subjects were shown the transcription of their oral tests two months after the test, and were asked to reflect on the language they produced and write down comments on their own speeches.

At the oral test, the subjects were paired randomly into eleven pairs, which were announced to them on the day of the test. The test was carried out in a small seminar room separately from other students. Upon entering the room, the subjects were given two tasks related to the topics and functions they had studied in class. For each task, they were given two minutes to talk, so each pair spoke for four minutes altogether. Six variations of tasks were prepared to be given to different pairs, so that later test takers would not get an advantage over the earlier test takers. The writer assessed the subjects according to the language produced (4 points), the interactive communication between the pairs (4 points), and the content of the conversation (4 points). The test was videotaped and transcribed by the writer. Examples of Task 1 and Task 2 are shown below:

Ex. Task 1 (variation 1)

<p>Student A: You are a CNN newscaster. Today you are going to interview a student from Jin-ai University. Ask the students about the following points: ・滞在目的 ・アメリカのいいところ ・足りないと思うこと ・日本とアメリカの違い</p>	<p>Student B: You are you. You are studying at the ALP (American Language Program) of California State University Fullerton, a sister-college of Jin-ai University. Today you are being interviewed by CNN.</p>
--	---

Task 2 (variation 1)

<p>Student A: You are you. You are visiting England for sightseeing. You have two weeks in London. You are being interviewed by BBC.</p>	<p>Student B: You are a BBC newscaster. You are reporting on foreigners sightseeing in London. You interview a foreign visitor (your partner) on a street in London. Ask him/her about the following points: ・出身地 ・滞在目的 ・ロンドンをどう思うか ・ロンドンと自分の出身地を比べてどう思うか</p>
--	--

3.3 Results:

3.3.1 Quantitative results:

Table 1 shows the quantitative results gained from the transcribed data and the questionnaire results. These results were used to see correlations between different variables. The correlations are shown in Table 2. Correlations were seen in three areas.⁸

First, MLU and the number of turns showed significant correlation. As the number of turns increases, MLU decreases. Since the length of the testing time is the same, the more turn-taking there was, the shorter they spoke in one turn. One interesting thing here is that

Table 1.

Subjects names	Proficiency levels	# of turns	MLU	gender	Test marks	acquaint1	acquaint2	difficulty	# of morphemes in one turn
S1	2	19	4.65	2	10	2	2	3	4.89
S2	2	19	4.77	2	10	1	1	3	6.52
S3	2	28	3.51	2	12	2	2	3	4.64
S4	2	27	4.71	2	12	2	2	3	6.63
S5	3	20	3.47	2	12	2	2	3	5.90
S6	2	20	4.09	2	12	2	2	3	6.55
S7	3	21	4.11	2	12	1	1	2	6.86
S8	3	21	5.04	2	12	1	1	3	6.24
S9	3	27	4.14	2	12	5	5	3	4.44
S10	3	28	3.76	2	12	3	5	4	5.64
S11	2	38	3.97	1	12	3	3	4	4.08
S12	1	38	2.89	2	11	3	3	4	3.34
S13	3	22	3.58	1	11	1	1	3	4.23
S14	3	22	3.63	2	12	2	1	3	5.27
S15	1	19	3.41	1	11	2	2	3	4.84
S16	3	19	4.67	2	12	1	1	3	5.16
S17	3	14	4.48	1	11	2	2	4	6.71
S18	2	14	4.32	2	12	2	2	4	5.86
S19	3	10	5.38	1	9	2	1	3	7.00
S20	2	10	5.47	2	11	2	2	4	10.40
S21	2	22	3.76	2	11	2	1	2	4.27
S22	2	22	2.96	1	9	2	1	2	3.50

- ◆ The proficiency levels (3=high, 2=middle, 1=low) refer to the results of a placement test the subjects had taken. The placement test had been devised by a private company to test the proficiency levels of the students' English in terms of reading, listening, vocabulary and grammar.
- ◆ The number of turns shows the number of turns in the testing time of 4 minutes in total. MLU (mean length of utterance) shows the average number of morphemes per T-Unit in the testing time of 4 minutes, repeated words not being counted.⁹
- ◆ Test marks are the test marks given by the teacher.
- ◆ 'Acquaint 1' shows the perceived acquaintanceship levels in the classroom about their interlocutors (1= have rarely spoken with each other ~ 5= have often spoken with each other).
- ◆ 'Acquaint 2' shows the perceived acquaintanceship outside classroom (1= have rarely spoken with each other ~ 5= have often spoken with each other).
- ◆ 'Difficulty' shows the perceived difficulty that the subjects felt about the test (graded from 1-5 with 5 being the most difficult).

the pairs who had more turn-takings gave the assessor an impression that they were more fluent than those who had fewer turn-takings. Although MLU is used to see the speakers' proficiency levels, it seems that there is a limitation in its use.

Secondly, the acquaintance level and the number of turns showed significant correlation. The more the subjects felt familiar with each other, the more they took turns. This is consistent with the findings of O'Sullivan (2000), which showed 'familiarity enhances test performance.'

Thirdly, the test marks and gender showed significant correlation. Female students got higher marks than male students. However, the validity of the assessors test rating was not tested, so the result can not be used to definitely decide on its statistical significance.

Table 2.

		Correlations							
		proficiency	turns	mlu	gender	testmark	acquain1	difficul	utterance
proficiency	Pearson Correlation	1	-.065	.211	.222	.366	-.140	-.058	.202
	Sig. (2-tailed)		.775	.347	.321	.094	.536	.798	.367
	N	22	22	22	22	22	22	22	22
turns	Pearson Correlation	-.065	1	-.583**	.085	.349	.449*	.119	-.654**
	Sig. (2-tailed)	.775		.004	.705	.112	.036	.598	.001
	N	22	22	22	22	22	22	22	22
mlu	Pearson Correlation	.211	-.583**	1	.142	-.097	-.241	.175	.754**
	Sig. (2-tailed)	.347	.004		.527	.668	.280	.436	.000
	N	22	22	22	22	22	22	22	22
gender	Pearson Correlation	.222	.085	.142	1	.492*	.032	-.030	.214
	Sig. (2-tailed)	.321	.705	.527		.020	.889	.896	.338
	N	22	22	22	22	22	22	22	22
testmark	Pearson Correlation	.366	.349	-.097	.492*	1	.147	.241	.041
	Sig. (2-tailed)	.094	.112	.668	.020		.515	.281	.855
	N	22	22	22	22	22	22	22	22
acquain1	Pearson Correlation	-.140	.449*	-.241	.032	.147	1	.320	-.281
	Sig. (2-tailed)	.536	.036	.280	.889	.515		.146	.205
	N	22	22	22	22	22	22	22	22
difficul	Pearson Correlation	-.058	.119	.175	-.030	.241	.320	1	.223
	Sig. (2-tailed)	.798	.598	.436	.896	.281	.146		.319
	N	22	22	22	22	22	22	22	22
utterance	Pearson Correlation	.202	-.654**	.754**	.214	.041	-.281	.223	1
	Sig. (2-tailed)	.367	.001	.000	.338	.855	.205	.319	
	N	22	22	22	22	22	22	22	22

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

3.3.2 Qualitative results:

This section will use the transcription of the language produced and the subjects' retrospective data to see the discourse features shown in the subjects' discourses.

1) IRF (initiate, reply, feedback) cycle: Most of the topic boundaries show the interviewers' initiating topics, followed by the interviewees' replies and then the interviewers' feedback on the replies. In other words, the discourse is quite reactive. Since there was no power relationship between the subjects, there was no asymmetrical feature that could exist in 'examiner-candidate' discourse. Only four discourse samples (Transcription 1(1), Transcription 9 (1), Transcription 10 (1) and Transcription 10 (2)) show 'initiate-reply'

patterns, lacking in 'feedback.' In Ex. 1, S1 is initiating questions, and S2 is responding. Without any feedback on the responses, the conversation goes on. So each boundary of one topic (1&2, 3&4, 5&6, 7&8) is short.

Ex. 1. from Transcription 1(1) (See Appendix)

- 1 S1: What (.) do you think (.) best (.) in America.
- 2 S2: Aaam. %What do you think best in America.% Eetto, I, I like the best in America, to gorgeous.
- 3 S1: What purpose is your visit.
- 4 S2: I'm studying at the American Language Program (.) of Fullerton College (.) a sister college of Jin-ai University.
- 5 S1: (3) What do you think (7) %chotto-matte%
- 6 S2: %n% (nod)
- 7 S1: (4) What (.) do you think, eee (3) difference of America (.) between and Japan.
- 8 S2: I: think, different, weather. Japa:n is humido weather, but America is dry weather.

On the other hand, when S2 becomes the interviewer, she is giving S1 feedback, so the boundaries of the topics become longer. In Ex. 2, from line 9 to line 14 is one boundary of one topic.

Ex. 2. from Transcription 1(2)

- 9 S2: Where are you from.
- 10 S1: Er I'm from (3) Jin-ai, ah I'm from in Japan.
- 11 S2: Japan. What's what's what's your prefecture.
- 12 S1: (4) Ah I'm live in Fukui.
- 13 S2: Fukui. Oh Fukui is many greens?
- 14 S1: Yes. There are a lot of green.

2) Collaboration toward successful communication: There is no example of collaborative completion, i.e., a pair's completing a sentence collaboratively. On the other hand, with fairly ungrammatical sentences, the subjects are responding with each other successfully. Ex. 3 shows S22 is guessing what S21 wanted to mean correctly, although S21 missed an important content word 'what' in line 15.

Ex. 3. from Transcription 11(1)

- 15 S21: Next, (3) do you think on learning English?
- 16 S22: Oh (.) n (.) (3) English (.) is (.) very difficult.
- 17 S21: Oh.
- 18 S22: Because I don't like English.
- 19 S21: Ooohh (.) you don't like English!

Collaborating with their interlocutors with description is also seen. In Ex. 4, S21 in line 24 wants to know what S22 in line 21 meant by saying America being very big. Then, S22 in

line 25 doesn't understand what S21 in line 24 wanted to know by saying 'why.' Now, from line 26, S21 starts helping S22 describe why America is big.

Ex. 4. from Transcription 11(1)

20 S21: Etto (.) first (.) what (.) do you think of America.

21 S22: What do you think. N. I think, ee (.) America is very, very, big.

22 S21: Oh (.) very big? (laugh)

23 S22: Yes. (laugh)

24 S21: Oh (.) why.

25 S22: hhh. why? (laugh) ee (laugh) why? Ee (.) nn (.) ee (.) nn (.) Ameri, n? (laugh)

26 S21: American scale?

27 S22: Oh (.) scale and land (.) land

28 S21: land

29 S22: land (.) big

30 S21: land (.) big (laugh)

Ex. 5. also shows how S9 and S10 are collaborating with each other in constructing sentences (meanings) in line 37 to line 44.

Ex. 5. from Transcription 5 (1)

31 S9: aa (2) What do you miss (1) in (1) your country.

32 S10: Ah. I miss my family. I was homesick. (laugh)

33 S9: Homesick. (laugh)

34 S10: But I'm OK,

35 S9: OK?

36 S10: Yes, because I have friends.

37 S9: Contact you family.

38 S10: Yes, maybe, once a week.

39 S9: Once a week. Write a letter?

40 S10 No.

41 S9: No?

42 S10: Call.

43 S9: Call? Too [expensive. (laugh)]

44 S10: [expensive. (laugh)]

3) Reduction and abandonment of communication goals: There are many cases of reduction of communication goals. Some interviewing students missed asking a question or two. This does not necessarily mean the abandonment of communication goals but just lack of time. In Ex. 6, the task given to S13 was to ask what S14 misses in America. Instead, S13 first tries to use 'lack' in line 45. Then finding that S14 doesn't understand his question, he says 'bad point' in line 47. The meaning S13 is creating here is quite far from the question 'what do you miss in America?' The reply he gets from S14 in line 52 does not adequately

answer the question that he should have asked.

Ex. 6. from Transcription 7 (1)

45 S13: nn. (4) Do you think lack about America.

46 S14: Lack?

47 S13: eee (3) eee (3) eee (5) bad point

48 S14: Good point

49 S13: about Americans.

50 S14: Good point?

51 S13: Bad point.

52 S14: Bad point. Erm (.) I think, American things very big, so, Japanese, for Japanese, too big,

- 4) "OK", "Next question", "OK, next", and gestures are used to change topics: Since the roles the subjects played were interviewers and interviewees, interviewers always took the leads in changing topics, by saying, 'OK,' 'Next question,' or by looking at the task sheet to see the next question to ask. This kind of action is also seen in official interview tests. In Ex. 7, S7 is constantly using 'OK' and 'next' and looking at the task sheet to change the topics (lines 53, 55, 59).

Ex. 7. from Transcription 4(1)

53 S7: OK. (1) Next (.) what do you (.) what do you like (1) America?

54 S8: nnnn I think this country is (.) ah (.) this country have many type of people. I can feel many culture, so this's my favourite point.

55 S7: OK. And (.) what do you spend time on your free time.

56 S8: nnn (.) now, I walk around this town.

57 S7: Where did you go?

58 S8: nnn. (2) I, go, I: went to the park yesterday.

59 S7: ern (.) OK. Next (.) what do you think, comparing America and Japan.

- 5) Repetition of the interlocutors' questions is often seen: In Ex. 8, S2 is repeating S1's question in a small voice looking into the air. It is as if she is processing the meaning aloud. Then she answers the question.

Ex. 8. from Transcription 1(1)

60 S1: What (.) do you think (.) best (.) in America.

61 S2: Aaam. %What do you think best in America.% Eeetto, I:, I like the best in America, to gorgeous.

- 6) Correction of misunderstanding (repair): Although the transcription didn't show so many apparent misunderstandings, Ex. 9 is an example of correction of misunderstanding. Line 63 shows S3 has misunderstood S4's question that she asked in line 62. Then S4 tries to correct S3's misunderstanding in line 64.

Ex. 9. from Transcription 2(2)

62 S4: (nod) n (.) what is (.) comparing (.) in London, and your town (.) ah (.) in London ando (.) Tokyo (.) in Tokyo.

63 S3: No. In Fukui.

64 S4: Oh(.) you live (.) you live in (.) OK (.) I want to ask what's comparing in London (.) comparing London and in Tokyo.

65 S3: Oh (.) I see.

7) Self-grammatical corrections (self-repair): The subjects are sometimes correcting their own sentences as they speak. In Ex. 10, S6 is focusing on the correct form of the verbs in line 66. However, her attempt at correction in line 68 fails, although the meaning is communicated to S5. S5 is also correcting herself in line 69, from 'more new' to 'newer.'

Ex. 10. from Transcription 3(2)

66 S6: (3) Do you, do you heard, did you hear, did you hear the bell?

67 S5: Yes.

68 S6: Yes. (5) Do you think, do you think about (3) London and your country (2) in comparing, compare,

69 S5: Yes. Ern (.) I think London is keep modern and old, but Japan is become more new, newer and newer, so I think London is (1) wonderful country.

8) Overlapping: Overlapping is seen very rarely. Compared to other transcriptions of conversations in English, this phenomenon of these Japanese subjects seems to be consistent with their conversation in Japanese.

9) No evidence is found to show that the acquaintanceship is causing problems in Ss' performances. Except S9 and S10, the subjects said they rarely had spoken with their partners before the test. This lack of partner familiarity does not seem to be hindering the collaborative conversation between the partners.

10) The number of turns and the number of morphemes in one turn: S11& S12 Pair (Transcription 6) and S19 & S20 Pair (Transcription 10) showed a unique contrast in terms of the numbers of turns and the numbers of morphemes in one turn. S11's average number of morphemes in one turn is 4.08, S12's 3.34, S19's 7.00, and S20's 10.40. Then, S11 and S12 had 38 turns each to speak, whereas S19 and S20 had only 10 turns. The quality of English language of S11 and S12 looks very simple, but the conversation flows fluently. On the other hand, although the MLU of S19 and S20 is comparatively high, 5.38 and 5.47 respectively, the conversation does not flow fluently, having a lot of long pauses.

11) Affective reaction to the test format: Affective reactions to the test format were seen in the subjects' descriptions in the questionnaire. Among negative reactions were; nervousness

(13/22 Ss), finding their interlocutor difficult to talk to (1/22 Ss), and not being happy being videotaped (2/22 Ss). Among favorable reactions were; "I had fun taking this test." (5/22 Ss), "It was fun, and I wished other tests were as enjoyable as this test." (1/22 Ss), "My partner tried to expand the topic" (1 /22 Ss), "I'm glad it was not in front of other students" (1/22 Ss), "I was anxious because I had not talked with the partner, but once the test started, he was easy to talk with, and I felt glad." (1/22 Ss)

12) The subjects' comments on their performances reading the transcription after two months of the test: These comments show the subjects are being able to look at their own language objectively and this retrospective process is encouraging 'noticing' about their 'interlanguage.'

S1: I was very nervous since we were the first test-takers. I had practiced, but I could hardly remember even simple words and questions. When I read the transcription, I was shocked. I could see how low my English level was. I feel sorry for my partner, too. I should have practiced harder. When I get another chance like this, I want to practice harder so as to come up with more words and have continued conversation. I will make more effort to speak English every day.

S3: I felt that there were many sentences unfinished. When I played the 'student' role, I noticed I was not able to give my opinions at all. I think I didn't know what to talk about rather than I didn't know how to say it in English. When I was the 'newscaster' , I was only asking questions and was not giving so much feedback to my partner. I want to be able to do this. I want to be able to speak English fluently.

S4: When I read this transcription, there were so many places that did not make sense and many ungrammatical sentences. I thought it was junior high school level. There were also misunderstandings, and I thought it was difficult to communicate something correctly. I was shocked since this was the first time I had seen the transcription of my own speech.

S5: When I was the 'interviewer', I was nervous and it took me a long time to talk, but when I was the 'interviewee', I was a bit more relaxed. I was surprised to know that I often used 'OK' to change the topics.

S6: I thought I was just saying words and not making sentences. There were many silences and unfinished sentences. I felt I couldn't come up with proper questions spontaneously. Both pronunciation and grammar are difficult in conversation.

S7: When I was asked of my opinions, I couldn't express myself well. In giving facts, I could speak without being at a loss.

S9: In this conversation, I was very nervous and I couldn't speak English well. Seeing this transcription two months after the test, I am surprised to see what I was saying.

S10: When I was the 'interviewer', I was in a panic and felt pressured to be humorous. When I was the 'student', I thought I had to reply well. When I read the transcription, I wondered why I couldn't speak better. I will try to give more thought to what I want to

say in English from now on.

S11: There was lack of vocabulary and the vocabulary I used was junior high school level. I spoke few words in one turn. I want to be able to speak more. I want to be able to say what I want to say more spontaneously.

S13: There were many silences, and the conversation was not smooth. I wish I could speak longer sentences.

S14: When I read the transcription, I thought I was not speaking well. The topics were not expanded so much. I thought I did well on the test day. I will try harder next time.

S16: I wondered how we could carry on the conversation like this. I thought I was saying 'eeto' too much unconsciously. I want to be able to speak more fluently.

S18: I remember I was very nervous. I was speaking only simple words. I want to study harder and speak about many things. When I was the 'interviewee', I could have talked more. I should have listened to my partner more carefully.

S20: When I read the transcription, I noticed I was laughing in many places. Probably I was too shy because I'm not used to speaking in English.

S21: I thought I was not carrying out the conversation well. When I was asked questions, I was answering with 'Oh' or 'Yes!', so now I feel sorry for my partner. I should have expanded the conversation more. My sentences were unstructured and I was just speaking word by word. I should have studied harder. It was good to have such a chance to speak in English.

4. Discussion:

This section will address the following two research questions based on the results gained in the previous section:

- (1) What could a paired oral test as an achievement test display?
- (2) What can be assessed by using paired oral tests as achievement tests?

The subjects were using various communication strategies, i.e., they made 'modifications to their intended goals during the planning process to compensate for gaps between what would be required to accomplish their intended goals and what they actually can do' (Yoshida, p.207). Among them were reduction strategies, achievement strategies, repair and fillers.

The basic characteristics of conversation summarised by van Lier (1989, p.495), 'face-to-face interaction, unplannedness, unpredictability of sequence and outcomes, potentially equal distribution of rights and duties in talk, and manifestation of features of reactive and mutual contingency,' were present, although questions were mainly provided by the interviewing students, who were mostly 'responsible for beginning and ending the interaction, for ending one topic and introducing a new topic.' This was because of the setting of the given situation, but this weakness was overcome by each subject's having to play both interviewer and interviewee roles.

The subjects were collaborating for successful outcomes. This would indicate that

the paired format would encourage collaboration among the subjects, thus encouraging negotiation of meanings for successful communication. In Canale and Swain's (1980) model of communicative competence, there are linguistic, sociolinguistic, discourse and strategic competences. By being paired, the subjects could show their discourse competence for coherent and fluid turn taking and strategic competence for successful negotiation of meaning in the case of potential communication breakdown. In other words, the paired-format could elicit discourse and strategic competences from the subjects.

No significant correlation was shown between 'proficiency' and the test marks. Although it has to be noted that the reliability of the rating scale was not tested here, we may say that achievement tests could assess students' efforts such as memorization of conversational patterns which may not have been acquired yet. Thus, it could show 'proficiency + α ' as is illustrated in Figure 1. The hypothesis here is that this '+ α ' could lead to the development of proficiency in the future after the language is acquired as a result of more learning.

The test marks show how the assessor, the teacher, judged the students' achievement levels. The reliability of this assessment was not verified here. At local school settings, the reliability of teachers' assessments could be a point of consideration and teacher training in this field should be necessary.

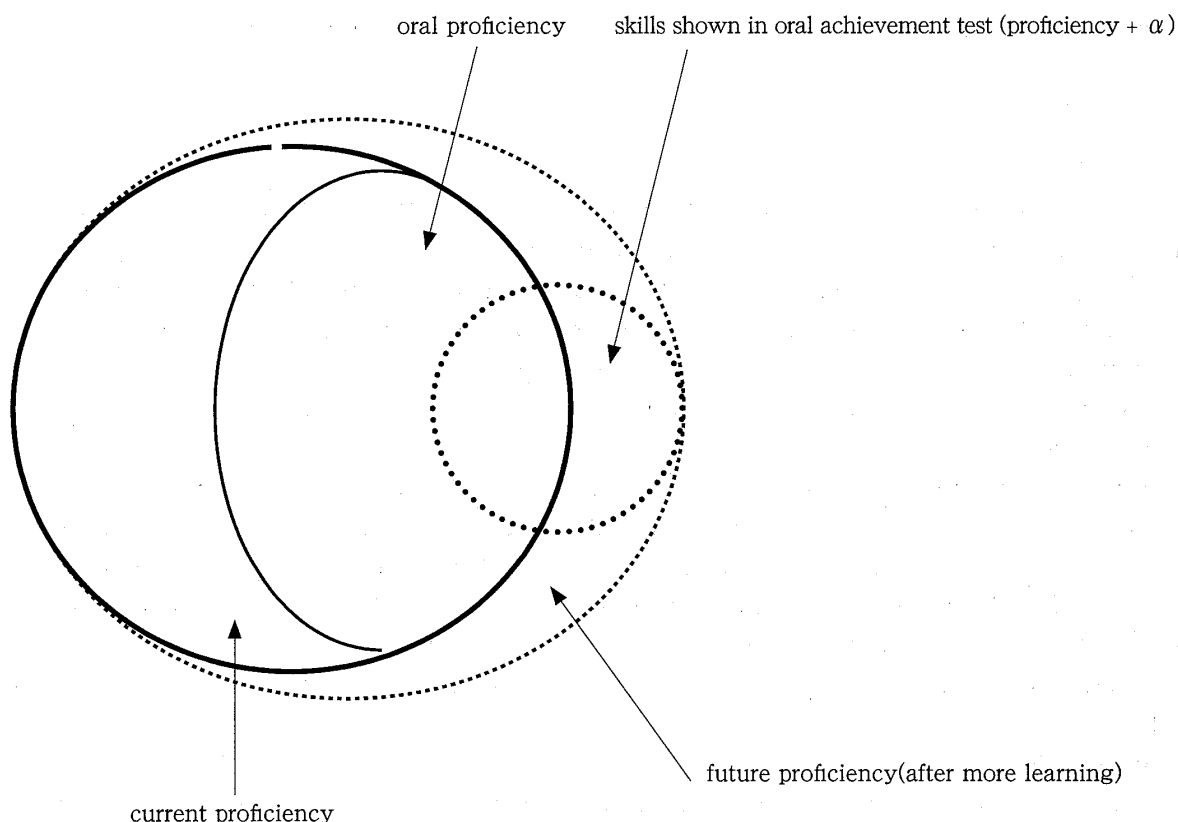


Figure 1. The relationship between the current proficiency and the skills shown in an oral achievement test

5. Conclusion:

This study carried out empirical research into the nature of the discourse displayed in a paired oral test. Various communication strategies were elicited in this format. It could solve the problem of an 'interview' test not realizing 'authentic conversation' (See van Lier: 1989, Lazaraton: 1992) as it could be said that it showed the validity of using paired oral tests as achievement tests. However, for a quantitative analysis, more data is needed. Differences in task types should be studied for this type of test to be introduced systematically into school practices.

In the development of oral tests as achievement tests at local schools, the following points would need to be considered:

- (1) The situation should be something students could imagine happening in real life situations.
- (2) The test has to be able to assess what students have studied in class. To create real-life like communicative situations, what to control and how to give freedom of speech need to be carefully considered.
- (3) Testing-time duration needs to be studied further in regard to its practicality and validity.
- (4) Transcription of the oral test could be done by students, too, as part of the feedback on the test. The data could also be filed as portfolios and used in final assessments. Transcription and analysis of the discourse could also give useful feedback on teaching.

Notes:

- ¹ The Course of Study is revised by the Japanese Ministry of Education every ten years. Schools are required to pursue the goals stated in the Course of Study.
- ² At senior high schools, "Oral Communication A", "Oral Communication B", "Oral Communication C" were taught between 1992 and 2002, and then the syllabus drastically being revised, "Oral Communication I" and "Oral Communication II" started in 2003.
- ³ Yamada(2004) reports 65% of her responding teachers in Fukui Prefectural senior high schools using interview tests.
- ⁴ Fulcher (2003) also writes about the situation in the U.K., and states that 'the introduction of the paired speaking test ... brings the test into line with classroom practice (p.186).' A paired format was introduced into 'the Certificate in Advanced English (CAE) in 1991, in the Key English Test (KET) in 1993 and for the revised version of the Preliminary English Test (PET) in 1995 and for revised First Certificate in English (FCE) in 1996(Langham, 2001).'
- ⁵ Egyud and Glover (2001) also discusses the advantages of having paired oral test at Hungarian schools.
- ⁶ See Galaczi's (2003).
- ⁷ Norton (2005) discusses Japanese female candidates' talking less when they are paired with male candidates.
- ⁸ The number of samples (N=22) is very small, so there is a limitation in this statistical analysis.
- ⁹ The use of T-Unit has a problem because it is basically used for the analysis of written texts. In oral texts, a combination of a 'subject' and a 'verb', which a T-Unit has to have, tends to collapse.
- ¹⁰ 'Transcription 1(1)' refers to the transcription of S1 and S2' s dialogue in carrying out Task (1).

So they are numbered from 'Transcription 1(1)' to 'Transcription 11(2)'. The whole transcription cannot be shown here for the matter of space.

References:

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1/1: 1-47.
- Egyud, G & Glover, P. (2001). Oral testing in pairs-A secondary school perspective. In *ELT Journal* 55/1: 30-37.
- Foot, M.C. (1999). Relaxing in pairs. In *ELT Journal* 53/1:36-41.
- Fulcher, G.. (2003). *Testing second language speaking*. Harlow: Pearson Education Limited.
- Galaczi, E.D. (2003). Interaction in a paired speaking test: the case of the First Certificate in English. *Research Notes* 14: 19-23.
- Ildiko, C. (2001). Is testing speaking in pairs disadvantageous for students? A quantitative study of partner effects on oral test scores. In *noVELty* 9/1.
- Lazaraton, A. (1992). The structural organization of a language interview: A conversation analytic perspective. In *System* 20/3: 373-386.
- Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge: Cambridge University Press.
- Lazaraton, A. (2006). Process and outcome in paired oral assessment. In *ELT Journal* 60/3: 287-289
- MEXT. (1998). "A vision of universities in the 21st century and reform measures" To be distinctive universities in a competitive environment. University Council Report.
Available at <http://www.mext.go.jp/english/news/1998/10/981010.htm>
- MEXT. (2003). Regarding the establishment of an action plan to cultivate "Japanese with English abilities." Available at http://www.mext.go.jp/b_menu/houdou/15/03/03033101/001.pdf
- MEXT. (2003). The course of study. Available at <http://www.mext.go.jp/english/shotou/030301.htm>
- Nakatsuhara, F. (2005). Impact of the proficiency-level on conversational styles in paired oral assessment. Retrieved from <http://www.iltaonline.com/newsletter/02-2005oct/org-baal.htm>. (2006/04/08)
- Norton, J. (2005). The paired format in the Cambridge Speaking Tests. In *ELT Journal* 59/4: 287-297.
- Norton, J. (2006). A response to Anne Lazaraton. In *ELT Journal* 60/3: 290-291.
- O'Sullivan. (2000). *Towards a model of performance in oral language testing*. Unpublished PhD. Reading: Reading University.
- Taylor, L. (2000). Investigating the paired speaking test format: Recent studies. In *Research Notes* 2: 14-15. Cambridge: UCLES.
- Taylor, L. (2001). The paired speaking test format: Recent studies. In *Research Notes* 6: 15-17. Cambridge:UCLES.
- Taylor, L. & Jones, N. (2001). Revising the IELTS Speaking Test. In *Research Notes* 4: 9-12. Cambridge: UCLES.
- van Lier. (1989). Reeling, writhing, drawling, stretching, and fainting in coils; Oral proficiency interviews as conversation. In *TESOL Quarterly* 23/3: 489-508.
- Yamada, H. (2004). Assessing student communication abilities in the oral communication subjects in Japanese local high schools: - Some suggestions for improvement -. In *Fukui-ken English Research Group Kaiho*. 92-114.
- Yoshida-Morise, Y. (1998). The use of communication strategies in language proficiency interviews. In Young, R. & He, A. W. (Eds.), *Talking and testing* (pp.205-238). Amsterdam: John Benjamins Publishing Co.

Appendix:

Transcription notation symbols (adapted from Lazaraton 2002):

1. unfilled pauses or gaps: Micropauses, those of less than .2 seconds are symbolized (.); longer pauses appear as a time within parentheses.
2. colon (:): a lengthened sound or syllable.
3. punctuation: markers of intonation rather than clausal structure; a period (.) is falling intonation, a question mark (?) is rising intonation, a comma (,) is continuing intonation.
4. brackets ([]): overlapping talk, where utterances start and/or end simultaneously.
5. percent signs (% %): quiet talk.
6. arrows (> <): the talk speeds up; arrows (< >): the talk slows down.
7. empty parentheses (): transcription doubt, uncertainty; words within parentheses are uncertain.

Acknowledgement

This research has already been the subject of two presentations. It was first presented at an informal forum at California State University Fullerton (CSUF) and subsequently at the 2006 Japan Association for Language Teaching (JALT) International Conference. The forum at CSUF was organized by the Departmental Chair of the Department of Modern Languages and Literatures, Dr. Janet Eyring. The forum took place on September 26th, 2006, while the writer was visiting CSUF with the students on the two-week ALP Program. The forum was an opportunity for the TESOL faculty members and the writer to exchange views on raising the oral proficiency of Japanese learners of English. Many kind suggestions were given in regard to the writer's current research and the writer would like to express heartfelt thanks to the TESOL faculty members and the ALP teachers at CSUF for their suggestions, although any errors in the paper are entirely hers. The writer hopes that this kind of academic exchange will continue and prevail among the staff of CSUF and Jin-ai University.