

Native Speakers' Assessment of Japanese Students' English Speaking Ability : A Pilot Study

Yuji Nakamura

1. Introduction

According to the research of the JACET study group (1989), most college students want to have oral/aural communication ability (especially speaking ability) as a result of their college English classes.

The "Gakushu-shido-yoryo" (Course of Study for Lower and Upper Secondary Schools in Japan) offered by the Mombusho (The Ministry of Education), which is going to go into effect in 1994, has advocated that students' communication ability should be enhanced to meet the needs of students and the society. The Mombusho has stressed the new aspects of the guidelines (oral/aural communication A, B, C) to support their claim of the needs of communication education.

Some textbooks, teaching techniques, materials for better communication are being developed, and some innovative teachers have already attacked the problem of the lack of students' communication ability by adopting novel devices.

Nevertheless, the testing side of communication ability (specifically speaking ability) has rarely been focused on in this communication boom. This is probably because testing of speaking ability has many problematic factors.

In fact, testing of speaking ability has more problems to be solved than the other three language skills (reading, writing and listening). Let us consider three crucial problems of speaking tests here at the moment.

The first problem is to establish the concept of speaking ability. There are two ways to look at the construct of speaking ability. On the one hand, existing speaking tests such as ILR (Interagency Language Roundtable, formerly FSI), TSE (Test of Spoken English), ACTFL (The American Council on the Teaching of Foreign Languages) Test, STEP (Society for Testing English Proficiency) Test have linguistic components in their scoring criteria :

ILR ((Accent (Pronunciation), Grammar, Vocabulary, Fluency, Comprehension))

TSE(Overall comprehensibility, Pronunciation, Grammar, Fluency)

ACTFL (Fluency, Grammar, Pragmatic competence, Pronunciation, Sociolinguistic competence, Vocabulary)

STEP (1st Grade : Contents, Delivery ; pre-1st Grade :

Pronunciation, Fluency, Vocabulary, Grammar, Contents ; 2nd Grade and 3rd Grade : Pronunciation, Stress, Rhythm, Speed, Accuracy of contents)

On the other hand, linguists have also attempted to arrive at a construct of speaking ability (communication ability in the broader sense):

Lado (1961) (pronunciation, stress, intonation, grammatical structure, vocabulary)

Harris (1969) (pronunciation, grammar, vocabulary, fluency, comprehension)

Hymes (1972) (Communicative Competence)

Canale and Swain (1980) (Grammatical competence, Sociolinguistic competence, Strategic competence)

Canale (1983) (Grammatical competence, Sociolinguistic competence, Discourse competence, Strategic competence)

Bachman (1990) (Grammatical competence, Textual competence, Illocutionary competence, Sociolinguistic competence, Strategic competence, Psychophysiological mechanisms)

The present researcher has conducted two preliminary studies to establish the components of speaking ability. The results are as follows :

Nakamura (1990) (Content area, Lexicographic area, Morphological and syntactical area, Delivery area, Discourse area)

Nakamura (1991) (Interactional factor (Illocutionary competence, Sociolinguistic competence, Strategic competence), Grammatical factor, Phonological factor, Vocabulary factor, Content factor, Fluency and Discourse factor, Confidence factor)

Notwithstanding all of these previous studies, a valid concept of speaking ability which can be applicable to the classroom speaking test has not been established yet.

The second problem is to choose the tasks to measure speaking ability. There are three types of tasks (tests); direct tests (interviews, role play), semi-direct tests (tape-recorded evaluation) and indirect tests (paper-pencil tests). Existing tests have been conducted differently. ILR (formerly FSI) is administered in the form of an interview, ACTFL in the form of an interview and a role play, TSE by the tape recording, STEP (EIKEN) in the form of a speech, material description, and question-answer activities. Each of these

tests has its own advantages and drawbacks. When teachers test the speaking ability of forty students in a classroom setting, the desired goal should be narrowed. Indirect tests, such as paper and pencil tests, are easy to administer and might have high correlation with other tests; however, they lack the real active production or performance which is crucially important in speaking tests. In other words, they have little face validity.

Direct tests involving face-to-face conversation with an interlocutor provide a high degree of face validity and content validity as means of general oral proficiency although direct interview tests in class are often considered less reliable than the more structured semi-direct tests.

Semi-direct tests are generally designed to yield total numerical scores reflecting the accumulation of discrete performance "points." Therefore, they are recommended for the diagnostic testing of particular linguistic features. Furthermore, they need less administration time and personnel economy than direct tests (Clark 1985).

However, the evaluation of the examinees' tape-recorded responses may be laborious and time-consuming while evaluation of examinees' performance in direct tests is usually carried out at the same time and require little additional time for evaluation.

Consequently, it is safe to say that both direct tests and semi-direct tests have advantages and drawbacks and that they are playing the complementary part of speaking tests.

A third problem is the scoring reliability. Since subjectivity is highly involved in the evaluation of speaking ability, it is inevitable to have some disagreement among raters. The more important

problem is that Japanese classroom teachers must use the same evaluation standard as native speakers ; otherwise, the best students rated by Japanese teachers of English only using their standard may not be able to communicate with native speakers. Therefore, the first step in this research is to increase the interrater reliability of native speakers to construct the standard or the criterion of native speakers which Japanese teachers can rely on.

2. Purpose

The present researcher has proposed a framework of speaking ability (Table 1) mainly based on Bachman's theoretical framework (1990), with some consideration of scoring criteria of ACTFL and TSE, and the present researcher's two previous studies. The first purpose of this research is to examine if this proposed framework is reasonable or not.

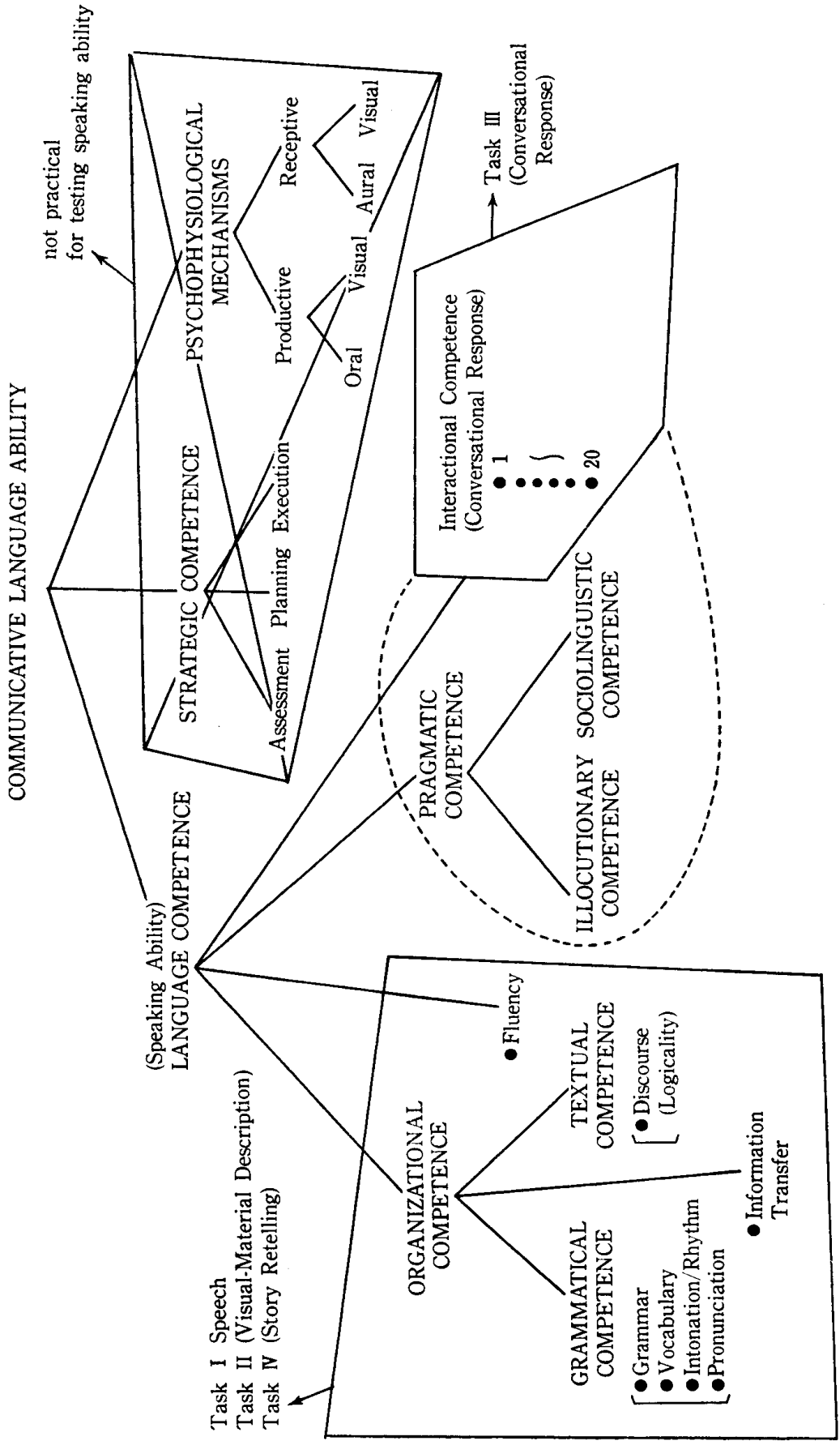
Secondly, the researcher has prepared four tasks shown below to test students' English speaking ability in the aforementioned proposed framework. The second purpose of this research is to see if these tasks are practical, authentic, and encourage students to speak.

Four Tasks Used for this Present Research :

1) Task 1 (Speech Making)

Students should choose one topic among eight topics, and prepare for a couple of minutes, then make speeches for three minutes. (Topics : My Friends, My Family, Part-time Work, My Hobbies, Traveling, Telephone Conversations, Fashion, College Life)

Table 1
 The Present Researcher's Proposed Framework of Speaking Ability



2) Task 2 (Visual-Material Description) (see Appendix 1)

Students should choose one item out of 21 choices and prepare for a couple of minutes and describe the item for three minutes (Choice Examples: Photograph, Map, Cartoon, Graph, Advertisement, Time Table, Itinerary)

3) Task 3 (Conversational Response Activities) (see Appendix 2)

Students should make responses to each question/sentence (twenty in total) recorded on tape. (Examples of Questions and Sentences: How are you?, What do you do?, What is the weather like today?, Will you do me a favor?, Do you mind if I use your eraser?)

4) Task 4 (Story Retelling) (see Appendix 3)

Students should read a short passage (approximately 100 words) for a couple of minutes and make a summary in their own words.

Thirdly, the present researcher has constructed a scoring sheet (Table 2 (1)–Table 2 (5)) which is deemed to cover the components of the speaking ability in the proposed framework. The third purpose of this research is to investigate if the scoring sheet is workable. In other words, the present researcher wants to examine if the scoring criteria in the scoring sheet can work well for the raters to separate the students into different labels.

Lastly, the present researcher wants to examine if the high inter-rater reliability is obtained or not among raters using the scoring sheets mentioned above.

Table 2 (1)**I. Speech**

Item		below average	average	above average	very good
1	Fluency				
2	Discourse (Logicality)				
3	Information Transfer				
4	Vocabulary (Appropriacy)				
5	Intonation and Rhythm				
6	Grammar (Accuracy)				
7	Pronunciation (Individual Sounds)				

Table 2 (2)**II. Visual Material Description**

Item		below average	average	above average	very good
8	Fluency				
9	Discourse (Logicality)				
10	Information Transfer				
11	Vocabulary (Appropriacy)				
12	Intonation and Rhythm				
13	Grammar (Accuracy)				
14	Pronunciation (Individual Sounds)				

Table 2 (3)

III. Conversational Response (See Appendix 2)

Item		no answer	conversationally inappropriate	conversationally appropriate	very good
Section 1					
15	No. 1				
16	2				
17	3				
18	4				
19	5				
20	6				
21	7				
22	8				
23	9				
24	10				
Section 2					
25	No. 1				
26	2				
27	3				
28	4				
29	5				
30	6				
31	7				
32	8				
33	9				
34	10				

Table 2 (4)

IV. Story Retelling

Item		below average	average	above average	very good
35	Fluency				
36	Discourse (Logicality)				
37	Information Transfer				
38	Vocabulary (Appropriacy)				
39	Intonation and Rhythm				
40	Grammar (Accuracy)				
41	Pronunciation (Individual Sounds)				
42	Understanding of Passage				

Table 2 (5)

V. Overall Impression

Item	below average	average	above average	very good
43				

3. Hypotheses

Hypothesis 1

The present researcher's proposed framework is realistic and reasonable.

Working Hypothesis :

There is no significant difference between the present researcher's proposed framework of speaking ability and the frame-

work or the criteria used by the native speakers.

Hypothesis 2

The interrater reliability is acceptable.

Working Hypothesis :

There is no significant difference of scoring reliability among raters.

Hypothesis 3

The four tasks (Speech Making, Visual–Material Description, Conversational Response Activities, Story Retelling) meet the standards set for them.

Hypothesis 4

The test was practical.

Working Hypothesis :

- 1) The length of the test did not cause test fatigue.
- 2) The time length of the test is reasonable.

4. Method

4.1. Subjects

- 1) Eighty–one college students majoring in English at Chofu College.
- 2) All of them are female and have been studying English in Japan for six years.
- 3) The range of their age is from eighteen to nineteen.

4.2. Tasks

Four tasks mentioned above (Speech making, Visual Material Description, Conversational Response Activities, Story Retelling) were conducted by those eighty–one students.

4.3. Evaluated Materials

Eighty-one recorded tapes in which students did their tasks in the language laboratory.

4.4. Raters

- 1) Four native speakers were chosen as raters.
- 2) They have been teaching English in Japan for at least one year.

4.5. Scoring Sheet (shown above)

Three tasks (speech making, visual-material description and story retelling) were rated on the 1-4 point scale (below average, average, above average and very good) in each linguistic component (vocabulary, pronunciation, grammar etc.). Conversational responses were rated on the 1-4 point scale (no answer, inappropriate, appropriate and very good).

4.6. Procedure

- 1) In the speech part, students chose one topic among eight choices and prepared for a couple of minutes and made speeches for three minutes.
- 2) In the visual-material description, students chose one item among twenty-one choices and prepared for a couple of minutes, then described the item for three minutes.
- 3) In the conversational-response part, students made responses to each question/sentence (twenty in total) recorded on tape.
- 4) In the story-retelling part, students read a short passage (100-150 words) and summarized it in their own words. They could take notes.

4.7. Data Analysis

- 1) Factor analysis was conducted to investigate how much the

researcher's framework agreed to that of native raters' evaluation or how much the raters' framework agreed to each other.

- 2) Interrater reliability was examined through Pearson's formula.
- 3) The U-L (upper-lower) analysis was conducted to examine what items can be useful to distinguish upper level students and lower level students.

5. Results and Discussion

5.1. Table 3 (Table 3 a (1)–Table 3 d (2)) shows the results of factor analysis of each rater.

Comparing the results of the factors in each rater (the results of Factor Analysis of the other raters were almost the same as Rater A's results), we can claim that there is only a minor difference in the definition of speaking ability among raters through the factors which were common to them. Moreover, if we further compare the results of four raters with the researcher's proposed framework of speaking ability, we should admit that there are similarities and differences between the two. The researcher's proposed framework consists of several factors such as (vocabulary, grammar, pronunciation, discourse, fluency, "interactional competence") rather independently, while four raters evaluated students' speaking ability globally in two ways; a) through "linguistic competence" (vocabulary, grammar, pronunciation, discourse and fluency etc.) holistically and b) through "interactional competence." Therefore, Hypothesis 1 was only partially supported.

Judging from this result, when raters are evaluating speaking

Tadle 3 a (1)
Results of Factor Analysis (of Rater A)

Item \ Factor	Factor 1	Factor 2
A 11 vocabulary	.84	—
A 43 overall	.84	.25
A 1 fluency	.84	—
A 4 vocabulary	.82	—
A 8 fluency	.80	—
A 5 intonation	.80	—
A 12 intonation	.79	—
A 14 pronunciation	.73	—
A 35 fluency	.73	.39
A 39 intonation	.73	.31
A 3 information	.73	—
A 10 information	.71	—
A 7 pronunciation	.71	—
A 9 discourse	.71	—
A 13 grammar	.69	—
A 36 discourse	.68	.38
A 2 discourse	.67	.32
A 6 grammar	.67	—
A 42 understanding	.65	.28
A 41 pronunciation	.65	.31
A 37 information	.64	.35
A 38 vocabulary	.62	.49
A 40 grammar	.57	.48
A 18 How are you?	.34	.25

Factor 1
("Linguistic Competence")
Factor 2
("Interactional Competence")

Eigenvalue
Factor 1 15.88
Factor 2 3.49

Pct. of Var.
Factor 1 36.9
Factor 2 8.1

Cum. Pct.
Factor 1 36.9
Factor 2 45.0

A 1 to A 7
(Speech)
A 8 to A 14
(V-M Description)
A 15 to A 34
(Conv. Response)
A 35 to A 42
(Story Retelling)
A 43
Overall

— = below .20

Table 3 a (2)

Item	Factor	Factor 1	Factor 2
A 34	How about...	.23	.71
A 22	What is the...	—	.64
A 33	Would you...	.22	.64
A 25	Thank you...	—	.61
A 28	It's a...	.22	.57
A 19	What do you...	—	.54
A 30	I'd like you...	.28	.53
A 23	What do you...	—	.52
A 32	Do you mind...	.23	.51
A 26	Will you do me...	—	.48
A 31	I'll see you...	.37	.46
A 29	Let's have a...	.40	.46
A 27	Say hello to...	.38	.44
A 17	Could you...	—	.38
A 24	How do you...	—	.37
A 15	How do you do?...	—	.33
A 21	What is the...	—	.33
A 16	What is your name?...	—	.28
A 20	Can you tell me...	—	.24

Table 3 b (1)
Results of Factor Analysis (of Rater B)

Item	Factor 1	Factor 2
B 43 overall	.87	—
B 1 fluency	.86	—
B 7 pronunciation	.85	—
B 5 intonation	.85	—
B 14 fluency	.82	—
B 11 vocabulary	.80	—
B 4 vocabulary	.80	—
B 8 fluency	.79	—
B 41 pronunciation	.78	.25
B 13 grammar	.76	—
B 6 grammar	.75	—
B 36 discourse	.75	.31
B 35 fluency	.72	.37
B 38 vocabulary	.71	.38
B 3 information	.70	—
B 12 intonation	.70	—
B 40 grammar	.69	.42
B 2 discourse	.69	—
B 39 intonation	.69	.46
B 37 information	.68	—
B 42 understanding	.66	—
B 9 discourse	.63	—
B 10 information	.49	—
B 21 What is the date...	.41	—
B 15 How do you do?	—	—
B 20 Can you tell me the...	—	—

Table 3 b (2)

Item	Factor 1	Factor 2
B 33 Would you...	—	.74
B 34 How about...	.23	.66
B 28 It's a...	—	.66
B 23 What do you...	—	.55
B 27 Say hello to...	.27	.53
B 30 I'd like you...	.29	.53
B 29 Let's have a...	.30	.52
B 32 Do you mind...	—	.51
B 31 I'll see you...	—	.49
B 22 What is the...	—	.48
B 19 What do you...	—	.44
B 17 Could you...	—	.41
B 26 Will you do...	.23	.39
B 25 Thank you...	—	.30
B 24 How do you...	—	.28
B 18 How are you...	—	.21

Eigenvalue

Factor 1 15.0

Factor 2 3.6

Pct. of var.

Factor 1 35.6

Factor 2 8.6

Cum. Pct.

Factor 1 35.6

Factor 2 44.2

B 1 to B 7

(Speech)

B 8 to B 14

(V-M Description)

B 15 to B 34

(Conv. Response)

B 35 to B 42

(Story Retelling)

B 43

(Overall)

ability, they are not focusing on the detailed components such as grammar or vocabulary individually. However, their total score (overall score) may be greatly affected by separate components. The “interactional competence” was separately evaluated from the “linguistic competence” in the native speaker’s evaluation, which should be carefully considered, because we usually take it for granted that the overall/global evaluation includes almost all the aspects of speaking ability.

Table 3c (1)
Results of Factor Analysis (of Rater C)

Item	Factor 1	Factor 2
C 43 overall	.83	.23
C 35 fluency	.80	—
C 36 discourse	.80	—
C 39 intonation	.79	.28
C 38 vocabulary	.76	.21
C 41 pronunciation	.76	—
C 1 fluency	.74	—
C 8 fluency	.73	—
C 37 information	.72	—
C 40 grammar	.70	—
C 11 vocabulary	.70	—
C 12 intonation	.69	—
C 14 pronunciation	.68	—
C 2 discourse	.68	—
C 10 information	.67	—
C 42 understanding	.67	—
C 4 vocabulary	.66	—
C 5 intonation	.66	—
C 6 grammar	.65	—
C 9 discourse	.65	—
C 3 information	.62	—
C 7 pronunciation	.60	.24
C 13 grammar	.56	—
C 26 Will you do me a favor?	.34	.34
C 23 What do you usually...	.32	.29
C 31 I'll see you...	.31	.30
C 16 What is your name?	.29	—

Table 3c (2)

Item	Factor 1	Factor 2
C 34 How about...	.25	.67
C 33 Would you...	—	.65
C 30 I'd like you...	—	.62
C 32 Do you mind...	.22	.54
C 28 It's a...	—	.51
C 22 What is the...	—	.50
C 18 How are you...	—	.49
C 25 Thank you...	—	.49
C 17 Could you...	—	.42
C 24 How do you...	—	.41
C 19 What do you...	—	.41
C 29 Let's have a...	.30	.35
C 27 Say hello to...	.27	.35
C 21 What is the...	—	.30
C 20 Can you tell me...	.26	.26
C 15 How do you do...	—	.24

Eigenvalue

Factor 1 13.7

Factor 2 3.0

Pct. of var.

Factor 1 31.9

Factor 2 7.1

Cum. Pct.

Factor 1 31.9

Factor 2 39.0

C 1 to C 7

(Speech)

C 8 to C 14

(V-M Description)

C 15 to C 34

(Conv. Response)

C 35 to C 42

(Story Retelling)

C 43

(Overall)

Although raters evaluate the "linguistic competence" holistically, the factor loading to each item in each task is separate. Therefore, the role of each task is necessary and important to evaluate sub-components of "linguistic competence." In other words, each task is indispensable and plays a complementary role in the evaluation of speaking ability.

Table 3d (1)
Results of Factor Analysis of Rater D

Item	Factor 1	Factor 2
D 43 overall	.81	.27
D 40 grammar	.78	—
D 39 intonation	.75	—
D 35 fluency	.75	.25
D 42 understanding	.74	—
D 14 pronunciation	.74	—
D 13 grammar	.73	—
D 38 vocabulary	.73	—
D 6 grammar	.71	—
D 36 discourse	.70	.21
D 41 pronunciation	.69	—
D 7 pronunciation	.68	—
D 12 intonation	.67	—
D 37 information	.66	—
D 4 vocabulary	.66	—
D 1 fluency	.65	.25
D 5 intonation	.64	—
D 2 discourse	.61	.28
D 8 fluency	.58	.22
D 3 information	.52	.31
D 11 vocabulary	.51	.26
D 9 discourse	.50	.30
D 10 information	.47	.33
D 21 What is the date...	—	—
D 20 Can you tell me...	—	—

Table 3 d (2)

Item	Factor 1	Factor 2
D 34 How about...	—	.67
D 33 Would you...	—	.60
D 30 I'd like you...	—	.58
D 32 Do you mind...	.21	.54
D 31 I'll see you...	.27	.51
D 17 Could you...	—	.45
D 27 Say hello to...	.20	.44
D 22 What is the...	—	.42
D 29 Let's have a...	—	.42
D 28 It's a...	.22	.42
D 26 Will you do me...	.28	.41
D 19 What do you...	—	.41
D 23 What do you...	—	.40
D 25 Thank you...	—	.37
D 15 How do you do?	—	.34
D 24 How do you...	—	.28
D 18 How are you?	.24	.26

Eigenvalue
 Factor 1 12.4
 Factor 2 2.9
 Pct. of var.
 Factor 1 29.5
 Factor 2 7.0
 Cum. Pct.
 Factor 1 29.5
 Factor 2 36.5

D 1 to D 7
 (Speech)
 D 8 to D 14
 (V-M Description)
 D 15 to D 34
 (Conv. Response)
 D 35 to D 42
 (Story Retelling)
 D 43
 (Overall)

5.2. Table 4 demonstrates the results of interrater reliability.

Table 4
Interrater Reliability among Four Raters
(Pearson Correlation Coefficients)

Rater	A	B	C	D
A	—			
B	.81	—		
C	.86	.84	—	
D	.83	.83	.87	—

The interrater reliability was reasonable for a pilot test, given the experimental nature of the speaking test. However, we need to improve it by the training of raters, the modification of the tasks, and the reorganization of the scoring sheet.

5.3. Table 5 indicates the results of the U-L (upper and lower) analysis of students.

Table 5
Results of U-L Analysis of Students
Rater A

Item		Mean		S. D.		t	p
		Group 1 (n=23)	Group 2 (n=22)	Group 1	Group 2		
17	Could you spell it please?	2.87	3.36	1.22	.95	1.51	ns
20	Can you tell me the time?	2.70	2.95	.77	.72	1.17	ns
21	What is the date today?	2.57	2.82	.79	.96	.97	ns

Rater B

Item		Mean		S. D.		t	p
		Group 1 (n=22)	Group 2 (n=25)	Group 1	Group 2		
15	How do you do ?	3.59	3.27	.73	.54	.69	ns
16	What is your name ?	4.00	4.00	—	—	—	ns
17	Could you spell it please ?	3.23	3.52	1.31	1.12	.83	ns
18	How are you ?	3.64	3.80	.90	.58	.75	ns
22	What is the weather like today ?	3.36	3.80	1.18	.50	1.69	ns

Rater C

Item		Mean		S. D.		t	p
		Group 1 (n=22)	Group 2 (n=23)	Group 1	Group 2		
15	How do you do ?	2.82	2.91	.50	.67	.54	ns
16	What is your name ?	3.05	3.00	.21	.00	1.02	ns
17	Could you spell it please ?	2.32	2.87	1.04	.82	1.98	ns
18	How are you ?	2.91	3.01	.53	.42	1.26	ns
21	What is the date today ?	2.34	2.74	.85	.75	1.57	ns
25	Thank you for everything.	2.55	2.96	.86	.48	2.00	ns

The results (such seven items as 15, 16, 17, 18, 21, 22 and 25 were pointed out as non-significant items at least by two raters) show that the upper group could not be differentiated from the lower group in these seven items in the task of Conversational Response

Rater D

Item		Mean		S. D.		t	p
		Group 1	Group 2	Group 1	Group 2		
16	What is your name?	3.00	3.00	—	—	—	ns
17	Could you spell it please?	2.36	2.70	.95	.88	1.22	ns
18	How are you?	2.91	3.31	.68	.34	1.38	ns
19	What do you do?	2.27	2.74	.88	.69	1.98	ns
21	What is the date today?	2.23	2.25	.81	.79	1.23	ns
22	What is the weather like today?	2.86	3.00	.47	.00	1.40	ns
24	How do you come to school?	2.41	2.78	.85	.42	1.87	ns
25	Thank you for everything.	2.55	2.91	.91	.52	1.68	ns

Activities, while thirteen items in this task and all the items in the other tasks (Speech Making, Visual-Material Description, Story Retelling) plus overall impression worked well for the purpose of separating the students into some groups. It is easy to surmise that those seven items (or conversational questions) did not help construct the demarcation between the upper and the lower group students. The items were so easy that almost all the students could answer them. Although these questions do not have any discriminating power, they are all warming up questions and also can make students have a feeling of accomplishment in the test situation especially among the lower group students. Therefore, they should be retained in the test.

5.4. Other findings through the researcher's observation.

- 1) Judging from the researcher's observation, students did very well in the speech section. On the other hand, they did poorly in the conversational response activities. More importantly, students did not understand the meaning of the story retelling task, thus it seems that this task was not appropriate to this test. If students are given much time to prepare, they will remember the story and recall the whole story. If they can take notes, they will copy the story and read it. The story retelling task was not familiar to students and was not relevant as a test.
- 2) The length of time for the test was reasonable and practical because students did not seem to get tired doing the tasks.
- 3) Raters did not have to listen to the whole part of speech, visual-material description and story retelling. They finished grading each task within two minutes.

One thing that could be observed through the scoring procedure was that one rater used only one part of the scale for some items, which is not helpful for the project. We must give detailed explanations and directions to raters for the revised test so that raters will use the scale efficiently and accurately to evaluate/separate students.

- 4) Most of the students had difficulty continuing to speak for three minutes in each task. Such tasks as speech, visual-material description and story retelling can be shorter.

Consequently, Hypotheses 3 and 4 were partially supported.

6. Conclusion

The following conclusions can be drawn :

- 1) The framework the present researcher proposed by taking into consideration theoretical backgrounds and two previous studies did not completely agree with the native speakers' real evaluation system. However, there were many overlapping parts between the two.
- 2) Raters (native speakers) assessed students' speaking ability in two ways :
 - (1) they holistically assessed the "linguistic competence" of speaking ability
 - (2) they assessed the "interactional competence" separately from the "linguistic competence"

In other words, these raters tend not to focus on the individual category such as grammar, vocabulary when they assess students' speaking ability.

- 3) The interrater reliability (over .81) was acceptable, though some raters behaved differently from the others in some scoring items. They were not using the scale efficiently in these cases, which should be noticed for the present researcher's further study.
- 4) The U-L (upper and lower) analysis could differentiate the upper group and the lower group in most of the items. In such three tasks as Speech Making, Visual-Material Description, Story Retelling all the items showed an discriminating power to separate the students. However, as we saw earlier, there are seven cases in the task of Conversational Response

Activities in which both the upper and the lower groups did well. There was no significant difference between them. Nevertheless, these easy items should be maintained because they can be warming up drills and also can give poor students a feeling of completion which is necessary even during the test situation.

- 5) The four tasks, as a whole, worked well to urge students to speak up in the test situation. However, students were not familiar with the story retelling task. Their performance was not satisfactory as a testing object from the viewpoint of speaking-test task. Therefore, it should be replaced, perhaps, by a question-answer task which can be a good starter for the conversational response in the future research.

For future research, the following three necessary tasks should be carried out :

- 1) Examine the validity of this speaking test by comparing the result of this test with that of other standardized tests (in other words, to check the concurrent validity).
- 2) Enhance the scorer reliability between Japanese teachers of English and native English speaking teachers as well as the reliability among native English speaking teachers, because the eventual goal is to have Japanese teachers of English utilize the standard of native speakers.
- 3) Separate the intermediate level students into some groups besides the top and the lower students so that relevant remedies can be given to an appropriate level group.

With all these changes, the next revised test will hopefully be a more valid, reliable and practical one.

NOTE

This paper is based on a presentation at the 31st Annual Convention of the Japan Association of College English Teachers held at Waseda University on September 9, 1992.

BIBLIOGRAPHY

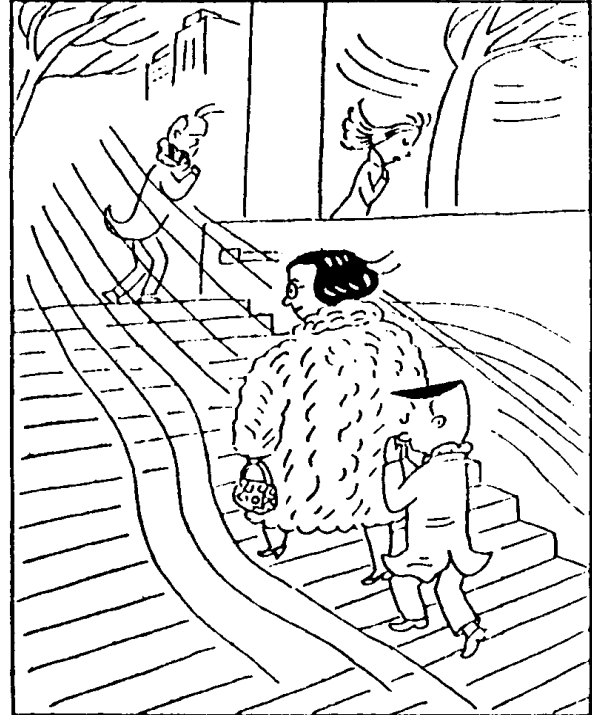
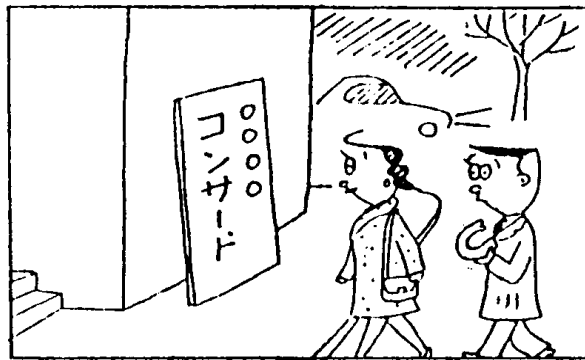
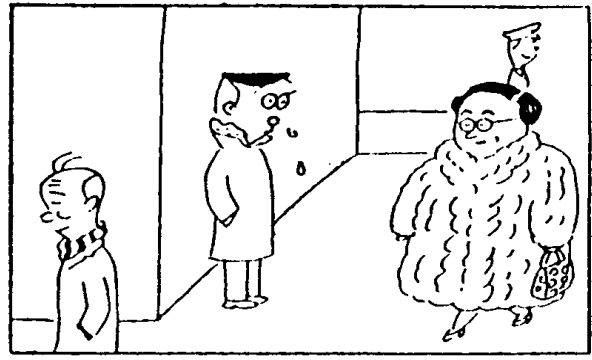
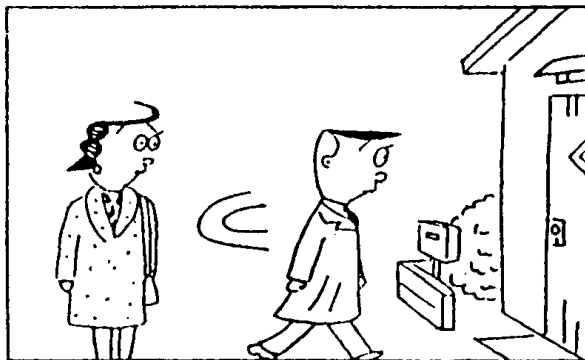
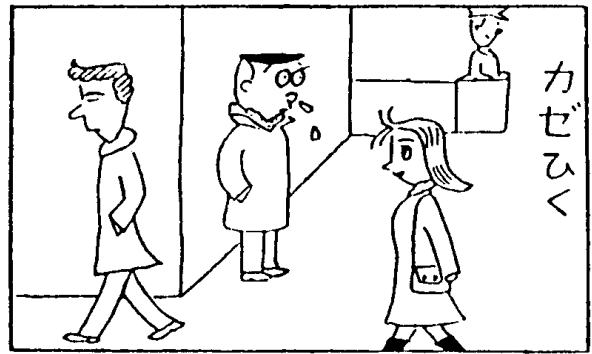
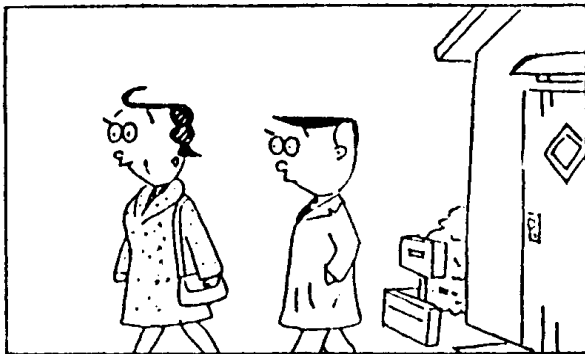
- ACTFL (The American Council on the Teaching of Foreign Languages). 1989. *The ACTFL Oral Proficiency Interview Tester Training Manual*. New York: The American Council on the Teaching of Foreign Languages.
- Bachman, L. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Canale, M. 1983. "On some dimensions of language proficiency" in J. W. Jr. Oller (ed.): *Issues in Language Testing Research*. Rowley, Mass.: Newbury House: 333-42.
- Canale, M. and M. Swain. 1980. "Theoretical bases of communicative approaches to second language teaching and testing." *Applied Linguistics* 1, 1: 47.
- Clark, J. L. D. 1985. "Foreign Language: Testing Speaking Proficiency" *The International Encyclopedia of Education: Research and Studies* 3: 1947-8.
- Harris, D. P. 1969. *Testing English as a Second Language*. New York: McGraw-Hill.
- Hymes, D. H. 1972. "On Communicative Competence" in J. B. Pride and J. Holmes (eds.) *Sociolinguistics*. Harmondsworth: Penguin: 269-93.
- ILR (Interagency Language Roundtable). 1987. "Interagency Language Roundtable Oral Proficiency Interview" in J. C. Alderson, K. J. Krahnke and C. W. Stansfield (eds.): *Reviews of English Language Proficiency Tests*. Washington D. C.: TESOL: 43-47.
- JACET. 1989. *A General Survey of English Language Teaching-College Graduates' Views*. Tokyo: The Japan Association of College English Teachers.
- Lado, R. 1961. *Language Testing*. New York: McGraw-Hill.

- Mombusho, The. 1988. *Course of Study for Lower and Upper Secondary Schools in Japan*. Tokyo : The Mombusho (The Ministry of Education).
- Nakamura, Y. 1990. "Varieties of Native Speakers' Evaluation on Non-Native Speakers' Spoken English" in *The Bulletin*. Tokyo : The Association of Private Jr. Colleges : 19 : 71-84.
- Nakamura, Y. 1991. "Differences in Native/Non-Native Teachers' Evaluation of Japanese Students' English Speaking Ability." Paper presented at the 30th Annual Convention of the Japan Association of College English Teachers, Hokkaido University, August 1991.
- Nakamura, Y. (in press). "Differences in Native/Non-Native Teachers' Evaluation of Japanese Students' English Speaking Ability." *Cross Currents* 19, 2.
- STEP (Society for Testing English Proficiency). 1991. *STEP Bulletin*. Tokyo : The Society for Testing English Proficiency : 3 : 132-33.
- TSE (Test of Spoken English). 1990. *TSE Manual for Score Users*. Princeton, N. J. : Educational Testing Service.

Appendix 1 (1)



Appendix 1 (2)



Appendix 1 (3)**Itinerary for Dr. Brown**

- July 20 (Sun) 21 : 10 Arrives at Narita from Singapore.
Stays at Narita Tokyo International Airport Rest House.
- July 21 (Mon) 10 : 00 Mr. Yoshida meets Dr. Brown at Narita and takes him to Tokyo.
Sightseeing in Tokyo.
17 : 00 Arrives at Keio Plaza Hotel.
Stays at Keio Plaza Hotel.
- July 22 (Tues) 9 : 00 Mr. Yoshida meets Dr. Brown and takes him to Tokyo Station.
10 : 30 Leaves Tokyo for Kyoto.
Stays in Kyoto Kokusai Hotel
- July 23 (Wed) 14 : 30 Arrives at Haneda by ANA 306 from Kyoto. Mr. Yoshida meets him at Haneda.
16 : 00 Arrives at Keio Plaza Hotel.
18 : 00 Dinner with Yoshida's Family in Shinjuku.
Stays at Keio Plaza Hotel.
- July 24 (Thur) 9 : 00 Leaves Keio Plaza Hotel for Narita.
14 : 00 Leaves Narita for London by British Airways.

Appendix 2

Section 1

Questions :

Item No.

- 15 1) How do you do?
- 16 2) What is your name?
- 17 3) Could you spell it please?
- 18 4) How are you?
- 19 5) What do you do?
- 20 6) Can you tell me the time?
- 21 7) What is the date today?
- 22 8) What is the weather like today?
- 23 9) What do you usually do on Sundays?
- 24 10) How do you come to school?

Section 2

Questions :

Item No.

- 25 1) Thank you for everything.
- 26 2) Will you do me a favor?
- 27 3) Say hello to your family?
- 28 4) It's a beautiful day, isn't it?
- 29 5) Let's have a cup of coffee.
- 30 6) I'd like you to meet my sister.
- 31 7) I'll see you at the restaurant at six tomorrow.
- 32 8) Do you mind if I use your eraser?
- 33 9) Would you like some ice cream for dessert?
- 34 10) How about playing tennis next Sunday?

Appendix 3

Story for Reading

An American lady took a present to her Japanese friend. She hoped that her friend would open it at once and be pleased with it. But after the Japanese friend thanked her for the present, she put it beside her and talked about other things. The American lady thought her friend was not interested in the present at all. She didn't know that it is not common in Japan to open any present in front of the person who gave it.