

Large Scale Measurement of English Speaking Ability

Yuji Nakamura

I. Theoretical background and rationale

Since the notion of Communicative Competence was publicized by Dell Hymes (1972) in his "On Communicative Competence," teachers, who had acknowledged the limitations of Structuralism and the limitations of the application of Transformational Generative Grammar to teaching, have been intrigued by the idea of Communicative Competence. The applied device of Communicative Competence called the "Communicative Approach" has come into vogue among English teachers around the world. Moreover, the communication boom has been the center of language teaching and learning along with the idea of the communicative approach in the field of language teaching and the notion of Communicative Competence (Canale and Swain, 1980; Littlewood, 1981; Richards and Schmidt, 1983). In other words, language teaching methodologies which concentrated on linguistic aspects of a language have been largely replaced by a new dimension of language teaching which focuses more on communication.

Performance testing (especially testing oral proficiency) has become one of the important issues in language testing since the role of speaking ability has become more central in language teaching along with the Communicative Approach. If we look at communication ability (especially speaking ability) in terms of testing, there is a great discrepancy between the expansion of the communication boom and the accurate measurement of communication ability (speaking ability) because of the difficulties involved in the construction and administration of a speaking test.

A review of the literature on speaking tests provides us with the following information of the actual situations, especially in Japan, of testing speaking ability (Nakamura, 1993; Shohamy, 1988; 1994).

- 1) Existing overseas speaking tests cannot be directly employed for assessing Japanese students' speaking ability in a classroom situation because of their level of difficulty and lack of practicality in the Japanese classroom situation.
- 2) There are almost no valid and reliable speaking tests available in a classroom setting to assess the lower and intermediate level students' speaking ability minutely.
- 3) Few studies are concerned with the definition of the construct of speaking ability.
- 4) Few Japanese scholars have conducted research on the influences of the speech modes (or the test tasks) on students' speaking performance.
- 5) There are no large scale speaking tests developed in Japan which are based on native speakers' scoring standards that can be easily conducted by Japanese teachers.
- 6) The need in Japan, especially at the college level, to improve English speaking ability and the testing of English speaking ability
- 7) The recognition of the problems of using productive speaking tests with students who have been accustomed to passive tests such as true false tests, or multiple choice tests and are hesitant to demonstrate their speaking ability in English
- 8) The ambiguity of the definition of speaking ability in the framework of Communicative Competence

II. Purpose of the research

The purpose of the research is to examine the detailed components of Japanese students' English speaking ability in terms of Communicative Competence by using the present researcher's constructed oral proficiency test which is theoretically based on Bachman's Communicative Language Ability model (1990). (See Table 1 in the Appendix). The test was examined to determine if it meets the following standards.

- 1) The construct validity is acceptable.
- 2) The concurrent validity is acceptable.

- 3) The washback validity is acceptable.
- 4) The face validity is acceptable.
- 5) The inter-rater reliability is acceptable.
- 6) The internal consistency reliability is acceptable.
- 7) Japanese teachers can use the test and rate Japanese students' English speaking ability without native speakers' assistance.
- 8) The test does not take an unduly long time to administer.
- 9) The four tasks are doable by Japanese college students.

III. Methods

Eighty college students took the test (see Table 2 in the Appendix) consisting of four tasks (Task I: Speech Making, Task II: Visual-Material Description, Task III: Conversational Response Activities, Task IV: Sociolinguistic Competence Test named Mini Contexts). (See Appendix for the test booklet).

Eleven raters* (4 Japanese and 7 native English speakers), who have been teaching English for at least one year, evaluated eighty audio tapes on which the students' responses had been recorded in the language laboratory. The raters used the scoring sheet and scoring criteria designed by the present researcher. The former two tasks were rated on a four point scale (below average, average, above average, very good) in each linguistic component such as grammar, vocabulary, pronunciation etc. Conversational responses were rated on a four point scale (no answer, conversationally inappropriate, conversationally appropriate, very good). Sociolinguistic competence answers were rated on a four point scale (no answer, sociolinguistically inappropriate, sociolinguistically appropriate, very good). (See Appendix for the scoring sheet and the scoring criteria).

* Eventually 10 raters' (4 Japanese and 6 Native) results were used for statistical analyses.

Data Analysis

- 1) Population Statistics

The mean and standard deviation of each rater's rating for each item were calculated to examine if each rater used the whole four point

scale of the scoring sheet.

2) Reliability

Each rater's raw score for each task was summed up to obtain a total score. Then, inter-rater reliability was measured through each rater's total score on the 80 tapes using Pearson's formula.

The internal consistency was examined through Cronbach's alpha to establish another measure of reliability.

3) Item Analysis

The item analysis (the investigation of item discrimination and item difficulty) was conducted in the following way.

(1) Item Discrimination

T-scores were used for item discrimination because the number of items for each task (six for Task I, six for Task II, 20 for Task III, 15 for Task IV) is not equal and a task with a greater number of items will tend to be weighted more heavily than the others.

(2) Item Difficulty

Item difficulty is usually reported as the percentage of successful students on each item. However, on a four point scale test such as the present study, the item mean and the standard deviation can be more appropriate ways to check the difficulty. Therefore, in the present item analysis study, the mean and the standard deviation of each rater in each item will be examined.

4) Correlation of the Four Tasks

Four tasks were examined from the viewpoint of the content validity. The four tasks should be mutually exclusive and only moderately inter-dependent to be composite of the whole proposed framework of speaking ability. Each task should represent each proposed competence (Linguistic Competence, Interactional Competence, Sociolinguistic Competence).

5) Factor Analysis

Factor analysis was conducted to examine the construct validity. Since the present researcher's proposed framework consists of three competences (Linguistic Competence, Interactional Competence,

Sociolinguistic Competence), this factor analysis will hopefully demonstrate if the proposed framework is reflected in the real evaluation situation.

6) Correlation between Four Tasks (by Raters) and Independent Variables

The concurrent validity was examined by looking at the correlation between four tasks (by raters) and teachers' class grades and a teacher's estimates.

7) Analysis of the Questionnaire Results of Three Native English Speaking Teachers Concerning their Grading System

The present researcher asked three native English speaking teachers at the researcher's college about the elements which make up the students' grades in their English conversation classes.

The question used for this questionnaire was: "What is included in student's grades of your English conversation class?" The present researcher also interviewed these three teachers about the details of their grading system.

8) Analysis of the Questionnaire Results from the Students Who Took the Speaking Test

The researcher asked those 80 students who took the speaking test to answer a questionnaire on their impressions of the test from the viewpoint of their study habits toward the improvement of their speaking ability.

The question used was, "Do you think this speaking test will change your study habits toward the improvement of your speaking ability?"

9) Tendency of Task Choice in the Speech Making Test and in the Visual-Material Description Test

In order to check if the task choice could be tightened by eliminating unnecessary or unpopular topics, the frequency of the topic selection was counted.

10) Examination of Practicality

The practicality and the usability of the test by Japanese teachers of

English will partly be demonstrated by showing the similarities between Japanese and native English speaking raters concerning 1) inter-rater reliability, 2) task correlation, and 3) factor analysis.

The test fatigue of the students and the length of the time of the test will be examined through direct observation.

IV. Results

1) Descriptive Statistics

The descriptive statistics (mean and standard deviation) (see Table 3 in the Appendix) indicate that we should have a comprehensive training session where a detailed explanation of the test purpose and the scoring procedure is given before the evaluation and that we should have at least two raters (if necessary three) evaluate each tape in order to have a reliable evaluation of students' speaking ability.

2) Inter-Rater Reliability and Internal Consistency Reliability

The inter-rater reliability (see Table 4 in the Appendix) was acceptable (over .74 among ten raters). In addition, a reasonably high correlation (the range is .74-.90 between individual native English speaking raters and Japanese raters) was obtained. This fact indicates that Japanese teachers by themselves can conduct the test and score the results in a classroom setting with little help or even with no help from native speakers (within the reliability range of .74-.90).

The internal consistency reliability (see Table 5 in the Appendix) (over .84 for 10 raters) was obtained and it showed that 47 items were measuring the students' speaking ability consistently.

3) Item Analysis (Item Discrimination and Item Difficulty)

Item Analysis was conducted to obtain item discrimination and item difficulty. Item Discrimination showed that 39 items out of the 47 in total were acceptable on statistical grounds. The remaining eight items (see Table 6 in the Appendix) did not have discriminating power, but they were retained for psychological reasons. In other words, most of the 47 items worked well to differentiate between the good group and the poor group.

Item Difficulty (see Table 7 in the Appendix) indicates that there is

almost no difference between the rating of Japanese and native English speaking raters.

One important discovery was that native English speaking raters were more severe about the level of Japanese students' fluency than Japanese raters. Another discovery was that the order of difficulty of the four tasks was found to be (easy to difficult) : first half of Task III, second half of Task III, Task IV, Task I, Task II. (See Table 8 in the Appendix).

4) Task Correlation (See Table 9 in the Appendix)

The results of task correlations demonstrate the strong correlations between Task I (Speech Making) and Task II (Visual-Material Description), and the tight correlation between Task III (Conversational Response Activities) and Task IV (Mini Contexts). The two pairs of tasks were playing a complementary role with each other with a strong relationship.

There was a high task correlation (over .81) between Japanese and native English speaking raters. In other words, Japanese raters will be able to evaluate students' speaking ability almost the same way as native English speaking raters. This task correlation also supported the construct validity in that the speaking ability consists of partially divisible four tasks.

5) Factor Analysis (See Table 10 in the Appendix)

Through factor analysis, we were able to obtain two factors: Factor 1 (Linguistic Ability) and Factor 2 (Interactional-Sociolinguistic Ability). (See Table 11 in the Appendix) Linguistic Ability is measured primarily by two tasks (Speech Making and Visual-Material Description) and Interactional-Sociolinguistic Ability is found in the other two tasks (Conversational Response Activities and Mini Contexts).

Both Speech Making and Visual-Material Description are scored using the same 6 criteria. But the level of difficulty of the two tasks was different. Conversational Response Activities have 20 discrete point items and the Mini Contexts subtest consists of 15 discrete point items.

Compared with Bachman's Communicative Language Ability model, the present researcher's final adjusted framework might be more practical and feasible because of the more concrete sub-components (items and situations) to cover the real and actual speaking ability than the theoretical model of

Communicative Competence.

Consequently, this two-factor structure, with help of task correlations, partially supported the present researcher's proposed framework of speaking ability based on Bachman's Communicative Language Ability model.

6) Concurrent Validity (See Table 12 in the Appendix)

Concurrent validity, one type of criterion related validity, was investigated by comparing the results of the present test with students' grades in English Conversation classes and a teacher's estimate of students' speaking ability.

The concurrent validity of this present test was supported not by the class grades but by the teacher's estimates. This is because students' grades include non-language proficiency elements such as attendance, effort and participation among other things.

7) Washback Validity

The washback validity was examined through the questionnaire from the students and the results supported the presupposition of washback validity. Students' responses suggested that they will change their study habits 1) by focusing more on the productive part of language skills, 2) by paying more attention to the contexts, etc. It is hoped that this change will reach teachers so that they will change their teaching styles from grammar-oriented to communication-oriented.

8) Face Validity

The face validity was partially supported by the present researcher's informal talk with students. They were excited about taking this type of unfamiliar but seemingly authentic speaking test. Therefore, they were highly motivated to speak out in the test situation.

9) Examination of Practicality

Practicality was already examined along with reliability and validity. However, there are some other aspects of practicality from the viewpoint of the test length and test booklet.

The present researcher's observation confirmed that there was no undue test fatigue for the students after the approximately 35 minute test.

The test booklet can be shortened by taking into consideration the results

of item choice analysis in Task I and Task II. In Task I (Speech Making) eight choices can become five, and in Task II (Visual-Material Description) 21 choices can become 10 for future research.

V. Implications of the results

The present speaking test based on Bachman's Communicative Language Ability model was able to shed light on the most undeveloped part of language skill testing — the measurement of oral proficiency — and was able to measure Japanese students' English speaking ability comprehensively in terms of Communicative Competence and helped us understand what speaking ability is.

This paper has shown that it is possible to construct a valid, reliable and practical test of speaking ability and shed light on native English speaking raters' scoring of Japanese students' speaking ability. It has also demonstrated that Japanese teachers will be able to conduct the test in a classroom setting by themselves easily, quickly, effectively and economically.

It is hoped that a wide adoption of this test will have a backwash/washback effect on the study habits of students and on the teaching, so that the whole tendency of English teaching in Japan will meet the students' needs of having communication ability (especially speaking ability).

For future research, the following points should be taken into consideration in order to try to make a more valid, reliable and practical test.

- 1) The test should be further examined in terms of further construct validation. (cf Shohamy 1994)
- 2) Other tests or estimates of speaking ability should be found so that criterion validity can be further checked. (cf Shohamy 1994)
- 3) The test should be administered to a much larger population of test takers including as many college students as possible to obtain more detailed information about Japanese students' speaking ability.
- 4) More examiners (especially Japanese raters) should be used so that the more detailed information can be obtained from less experienced teachers. (cf Douglas 1994)
- 5) The scoring criteria of this test should be compared to those of other

tests including teacher-made tests.

Note

This paper is based on my doctoral dissertation submitted to International Christian University 1993. Also a version of this article was presented at the 13th ACROLT (Academic Committee for Research on Language Testing) Language Testing Symposium Kiryat Anavim, Israel 1994, at the 177th Monthly Meeting of the JACET (Japan Association of College English Teachers) Tokyo 1994, and at the 19th Annual Congress of the Applied Linguistics Association of Australia Melbourne, Australia 1994.

References

- Bachman, L. F. (1990) : *Fundamental Considerations in Language Testing*. Oxford : Oxford University Press.
- Canale, M. and Swain, M. 1980 : Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1-47.
- Douglas, D. 1994 : Quantity and quality in speaking test performance. *Language Testing* 11, 125-144.
- Hymes, D. H. 1972 : On communicative competence. In Pride, J. B. and Holmes, J., editors, *Sociolinguistics*, Harmondsworth : Penguin.
- Nakamura, Y. 1992 : Differences in native/non-native teachers' evaluation of Japanese students' English speaking ability. *Cross Currents* 19, 2, 1-5.
- Nakamura, Y. 1993 : Measurement of Japanese college students' English speaking ability in a classroom setting. Unpublished PhD dissertation, International Christian University.
- Littlewood, W. 1981 : *Communicative Language Teaching*. London : Cambridge University Press.
- Richards, J. C. and Schmidt, R. W. 1983 : *Language and Communication*. New York and London : Longman.
- Shohamy, E. 1988 : A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition* 10, 165-79.
- Shohamy, E. 1994 : The validity of direct versus semi-direct oral tests. *Language Testing* 11, 99-123.

APPENDIX I

Table 1

Comparison between the Bachman model and the Present Researcher's Framework

Bachman	Present Researcher
Communicative Language Ability :	-----
Language Competence :	Speaking Ability
1) Organizational Competence :	1) Linguistic Competence
(1) Grammatical Competence	-----
(2) Textual Competence	-----
2) Pragmatic Competence :	2) Interactional Competence
(1) Illocutionary Competence	-----
(2) Sociolinguistic Competence	-----
3) ----- :	3) Sociolinguistic Competence
Strategic Competence :	***
Assessment	
Planning	
Execution	
Psychophysiological Mechanisms :	***
Productive	
Oral	
Visual	
Receptive	
Oral	
Visual	

N. B.

- 1) While there may be no exact one to one correspondence between Bachman's terms and the present researcher's terms, the nature of the test itself involves integration of the aspects of speaking ability. Consequently, "grammatical competence," "textual competence" or "illocutionary competence" are at least indirectly incorporated into certain of the items.
- 2) The asterisk (*) means that the category is not practical for assessing speaking ability.

Table 2

The Present Researcher's Theoretical Framework of Speaking Ability for the Present Test

Speaking Ability	
“Linguistic Competence”	“Interactional Competence”
Task 1 (Speech Making)	Task 3 (Conversational Response Activities)
1) fluency	20 sentences or
2) discourse	questions
3) vocabulary	
4) grammar	
5) pronunciation	
6) content	
Task 2 (Visual-Material Description)	Task 4 (Mini Contexts)
1) fluency	15 mini contexts
2) discourse	
3) vocabulary	
4) grammar	
5) pronunciation	
6) content	

Table 3
Mean, SD of Each Rater in Each Task

Rater	Task 1		Task 2		Task 3		Task 4	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
A	1.85	.78	1.97	.74	2.87	.42	2.56	.40
B	2.32	.80	2.15	.71	3.06	.38	2.47	.37
C	1.81	.65	1.70	.60	2.75	.41	2.31	.38
D	1.83	.70	1.60	.63	2.66	.32	2.32	.31
E	1.60	.52	1.56	.46	2.72	.37	2.27	.32
F	2.06	.64	1.97	.58	2.71	.33	2.50	.35
G	2.08	.60	1.83	.42	2.96	.38	2.34	.30
H	2.07	.65	1.93	.60	2.91	.40	2.38	.30
I	2.11	.45	1.96	.44	2.91	.39	2.22	.39
J	1.68	.42	1.65	.41	2.52	.24	2.11	.12
K	1.97	.44	1.93	.44	2.89	.37	2.42	.26

Table 4
Results of Inter-Rater Reliability
(Pearson's Correlation Coefficients)

	A	B	C	D	E	F	G	H	I	J	K
A											
B	.88										
C	.79	.83									
D	.91	.83	.76								
E	.90	.86	.84	.87							
F	.90	.88	.80	.89	.87						
G	.76	.74	.80	.77	.85	.77					
H	.85	.86	.78	.82	.91	.88	.87				
I	.84	.86	.85	.84	.91	.85	.88	.91			
J	.75	.77	.66	.74	.74	.73	.69	.80	.77		
K	.85	.86	.76	.84	.90	.83	.81	.87	.85	.72	

N. B. A-D: Japanese Raters

n=80

E-K: Native English Speaking Raters

All coefficients were significant at the .001 level.

Table 5
Internal Consistency Reliability

rater	alpha
A	.95
B	.92
C	.94
D	.94
E	.93
F	.92
G	.84
H	.91
I	.90
K	.91

Table 6
Content of the remaining eight items

Item	Task	Description
13	III	Nice to meet you.
14	III	What is your name?
16	III	How are you?
22	III	How do you come to school?
26	III	It's a beautiful day, isn't it?
31	III	Would you like some ice cream for dessert?
35	IV	Asking for repetition
46	IV	Offering

Table 7
Item Difficulty

item		J		NJ		ALL	
		M	SD	M	SD	M	SD
1	(1)	2.07	.62	2.15	.51	2.12	.51
2	(2)	1.95	.65	1.89	.48	1.91	.49
3	(3)	1.92	.66	2.02	.47	1.98	.51
4	(4)	1.92	.68	2.04	.51	1.99	.54
5	(5)	2.03	.75	1.84	.56	1.92	.60
6	(6)	1.82	.66	1.95	.51	1.90	.54
7	(1)	2.03	.58	1.99	.47	2.00	.48
8	(2)	1.74	.55	1.77	.42	1.76	.43
9	(3)	1.82	.57	1.92	.39	1.88	.42
10	(4)	1.88	.64	1.96	.49	1.93	.51
11	(5)	1.92	.70	1.72	.55	1.80	.58
12	(6)	1.75	.64	1.81	.46	1.79	.48
13	(1)	3.64	.55	3.57	.49	3.60	.50
14	(2)	3.88	.34	3.74	.25	3.80	.28
15	(3)	3.16	.49	3.18	.52	3.17	.49
16	(4)	3.69	.36	3.63	.31	3.65	.28
17	(5)	3.01	.83	3.15	.75	3.09	.76
18	(6)	3.08	.68	3.06	.63	3.07	.63
19	(7)	2.84	.78	2.87	.68	2.85	.70
20	(8)	3.18	.61	3.09	.59	3.12	.57
21	(9)	2.83	.61	2.85	.57	2.84	.57
22	(10)	2.96	.62	2.97	.46	2.97	.50
23	(11)	3.01	.97	2.98	.96	2.99	.95
24	(12)	1.72	.97	1.85	1.06	1.80	1.01
25	(13)	2.09	.90	2.08	.89	2.08	.88
26	(14)	3.03	.55	3.10	.46	3.07	.47
27	(15)	2.30	.83	2.42	.81	2.37	.80
28	(16)	2.00	.87	2.02	.88	2.01	.86
29	(17)	2.53	.89	2.50	.87	2.51	.87
30	(18)	2.04	.84	2.30	.97	2.20	.90
31	(19)	2.99	.49	3.03	.48	3.01	.45
32	(20)	2.69	.78	2.66	.73	2.67	.72
33	(1)	2.69	.61	2.57	.54	2.62	.54
34	(2)	2.55	.52	2.45	.48	2.49	.47
35	(3)	2.59	.49	2.56	.43	2.57	.42
36	(4)	2.23	.50	2.40	.54	2.33	.49
37	(5)	2.20	.59	2.24	.50	2.22	.51
38	(6)	2.46	.45	2.45	.50	2.45	.45
39	(7)	2.28	.48	2.18	.37	2.22	.39
40	(8)	2.40	.40	2.14	.43	2.45	.39
41	(9)	2.26	.56	2.18	.48	2.21	.48
42	(10)	2.21	.50	2.13	.43	2.16	.43
43	(11)	2.60	.68	2.54	.63	2.56	.63
44	(12)	2.60	.44	2.38	.44	2.47	.41
45	(13)	2.28	.62	2.14	.51	2.20	.52
46	(14)	2.75	.62	2.78	.56	2.77	.57
47	(15)	2.14	.74	2.20	.67	2.17	.68

J : Japanese raters as one group

NJ : Native English speaking raters as one group

ALL : Combination of J and NJ as one group

* Numbers in the parentheses are item numbers in each task.

Table 8
Task Difficulty

Task	J		NJ		ALL(J+NJ)	
	M	SD	M	SD	M	SD
1	1.95	.63	1.98	.46	1.97	.52
2	1.86	.57	1.86	.40	1.86	.46
3	2.83	.37	2.85	.35	2.84	.36
3-1	3.23	.32	3.21	.30	3.22	.31
3-2	2.44	.50	2.49	.47	2.47	.48
4	2.42	.33	2.35	.29	2.39	.30

N. B. 3-1 : First half of Task 3
3-2 : Second half of Task 3

Table 9
Inter-Task Correlations of Ten Raters as One Group
(Using T-scores)

	Task 1	Task 2	Task 3	Task 4	All
Task 1					
Task 2	.78				
Task 3	.57	.65			
Task 4	.66	.68	.78		
All	.85	.88	.87	.90	

N. B. Each correlation was significant at the .001 level of significance.

Table 10
Factor Analysis (2 Factors)
(ten raters as one group)

<i>item</i>	<i>task</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Communality</i>
5	1	.82	.16	.52
3	1	.80	.16	.56
6	1	.79	.14	.66
4	1	.79	.14	.64
2	1	.74	.10	.69
11	2	.73	.23	.65
8	2	.71	.19	.53
9	2	.71	.24	.54
1	1	.71	.17	.56
12	2	.70	.23	.44
7	2	.69	.23	.59
10	2	.62	.25	.54
37	4	.34	.34	.15
31	3	.25	.09	.10
25	3	.20	.58	.32
15	3	.04	.56	.08
45	4	.20	.53	.21
21	3	.15	.51	.18
23	3	.04	.49	.23
41	4	.27	.48	.25
19	3	.08	.47	.28
20	3	.16	.47	.15
43	4	.18	.47	.24
28	3	.34	.45	.21
17	3	.10	.45	.38
42	4	.24	.45	.18
44	4	.20	.44	.23
38	4	.18	.44	.32
29	3	.20	.44	.23
46	4	.02	.43	.18
27	3	.22	.43	.07
33	4	.33	.41	.18
34	4	.28	.40	.28
24	3	.23	.40	.24
13	3	.01	.39	.11
36	4	.13	.39	.17
47	4	.21	.38	.23
32	3	.21	.38	.22
22	3	.08	.38	.20
26	3	.22	.36	.16
30	3	.24	.35	.30
18	3	.28	.33	.26
40	4	.23	.33	.25
39	4	.30	.32	.23
35	4	.17	.29	.32
16	3	.11	.27	.19
14	3	-.03	.25	.19
Eigen.		11.4	3.0	
Pct.of		24.3	6.3	
Var.				
Cum. Pct		24.3	30.6	

Table 11
Revised Framework of Speaking Ability with Tasks

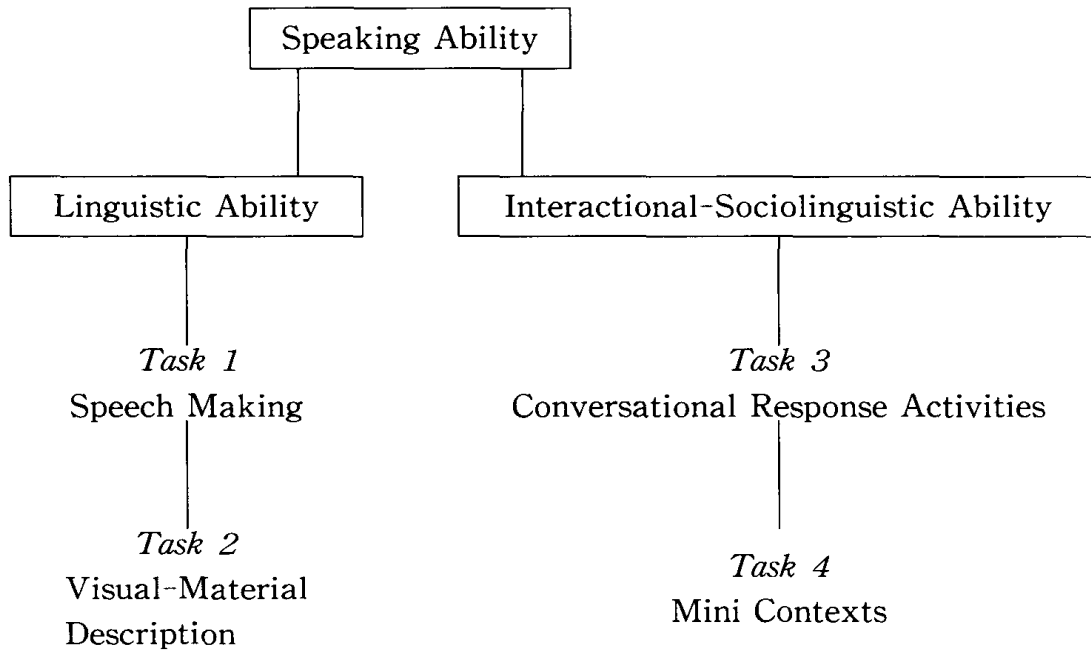


Table 12
Correlation between Teacher's Estimate and Four Tasks
(Pearson's Correlation Coefficients)

Task	TKA	TKB	TA	TKA : Conversation Class A (Grades) TKB : Conversation Class B (Grades) TA : Teacher's Estimate
1	.09	.17	.71	
2	.11	.10	.70	
3	.17	.13	.70	
4	.10	.16	.70	
Total	.14	.16	.80	

APPENDIX IITest Booklet, Scoring Sheet and Scoring Criteria

1). Speech (Two-minute Speech) Making Test

Directions :

- 1) Please choose one topic you want to talk about from among the eight topics given below.
- 2) Please take five minutes to prepare your speech.
- 3) Please give a two-minute speech about the topic you have chosen by giving the reason you chose it.

Topics :

- | | | |
|----------------------------|-----------------|-------------------|
| 1) My Friends | 2) My Family | 3) Part-time Work |
| 4) My Hobbies | 5) Traveling | 6) Fashion |
| 7) Telephone Conversations | 8) College Life | |

2). Visual-Material Description Test

Directions :

- 1) Please choose one item from among the following you would like to describe.
- 2) Please take five minutes to prepare your description.
- 3) Please describe the item you chose or give as much information as possible about it within two minutes.

N. B. Some of the examples are a picture, a chart, a map, a TV program, a cartoon or an itinerary.

3). Conversational Response Test

Directions :

- 1) You will hear twenty questions or sentences in English each followed by a pause.
- 2) Please give a quick and appropriate response in English to each sentence.

Twenty recorded sentences or questions :

- 1) Nice to meet you.
- 2) What is your name?
- 3) Could you spell it please?
- 4) How are you?
- 5) What do you do?
- 6) Can you tell me the time?
- 7) What is the date today?

- 8) What is the weather like today?
- 9) What do you usually do on Sundays?
- 10) How do you come to school?
- 11) Thank you for everything.
- 12) Will you do me a favor?
- 13) Say hello to your family.
- 14) It's a beautiful day, isn't it?
- 15) Let's have a cup of coffee.
- 16) I'd like you to meet my sister.
- 17) I'll see you at the restaurant at six tomorrow.
- 18) Do you mind if I use your eraser?
- 19) Would you like some ice cream for dessert?
- 20) How about playing tennis next Sunday?

4) . Sociolinguistic (Mini Contexts) Competence Test

Directions :

- 1) You will hear fifteen contexts in Japanese each followed by a pause
- 2) Please give an appropriate response in English in each context.

Context 1 (Apologizing and making an excuse)

You are late for your class. You missed the school bus.

Please apologize and make an excuse to your teacher.

Context 2 (Complaining and requesting)

You are in a non-smoking section of a waiting room at the airport. Someone started smoking. You have a cold and a sore throat. Please complain about it and request him/her to stop it.

Context 3 (Asking for repetition)

You didn't understand what your teacher said. You want the teacher to repeat it. Please make a request to your teacher.

Context 4 (Questioning)

You want to know the train schedule. Please ask about the departure time of the next train for Kyoto at the ticket office.

Context 5 (Greeting)

You happen to meet your high school teacher (Mr. Suzuki) after a long interval. Please greet him.

Context 6 (Parting)

After talking a while, you part from your teacher. Please say "farewell" to him.

Context 7 (Disagreeing)

Your friend (Tomoko) says jogging is a healthy activity. You don't agree with her. What do you say to her?

Context 8 (Congratulating)

Your friend's older sister won the first prize in a speech contest. Please congratulate her on her success.

Context 9 (Interrupting)

Your supervisor is working in his office. You want to interrupt him for a moment to talk with him. What do you say?

Context 10 (Warning)

Some children are playing baseball and almost break the window of your house. Please warn them.

Context 11 (Telephoning)

You are making a phone call. You want to speak to Mr. Brown. What do you say?

Context 12 (Telephoning)

You answer the phone. Someone wants to talk with your father. But he is out now. What do you say?

Context 13 (Getting an opinion)

You want to get your friend's opinion about last week's college festival. What do you say?

Context 14 (Offering)

You want to serve something to drink to a guest at your house. Please offer something to drink.

Context 15 (Asking for information)

At a department store, please ask the receptionist where the stationery section is.

Content of Scoring Items for Speech Test and Visual Material Description Test

pronunciation

- 1) segmental features (individual sounds ; vowels and consonants)
- 2) suprasegmental features (stress, rhythm, intonation)
- 3) enough volume (audible or not)
- 4) proper tone of voice

grammar

- 1) tense and aspect

- 2) noun-verb agreement
- 3) word order
- 4) noun-personal pronoun agreement

vocabulary

- 1) variety of words
- 2) word choice
- 3) idioms

content

- 1) creativity
- 2) novelty of topics

fluency

- 1) ease of speaking
- 2) speed of speaking
- 3) length, place and frequency of pauses

discourse

- 1) cohesiveness
- 2) logical combination of sentences

N. B.

All of these detailed sub-categories are employed from the results of Nakamura's research through a questionnaire given to 76 college English teachers (32 Japanese and 44 native English speaking teachers) (Nakamura 1991a ; Nakamura 1992a).

Criteria for Four Labels in Speech Test and Visual Material Description Test

below average

Chofu students who are below average (25%)

average

Average Chofu students (25%)

above average

Chofu students who are above average (25%)

very good

Outstanding Chofu students (25%)

Criteria for Four Labels in Conversational Response Test

no answer

- there is no response
- no answer
- meaningless sounds
- incomplete phrases
- incomprehensible words/phrases

conversationally inappropriate

- the answer is conversationally inappropriate
- conversationally not acceptable
(grammatically correct or grammatically incorrect)

conversationally appropriate

- the answer is conversationally appropriate
- conversationally acceptable
(grammatically not perfect but acceptable in terms of communication)

very good

- the answer is conversationally appropriate and also very good
- grammatically correct and conversationally appropriate
- made with ease

Criteria for Four Labels in Sociolinguistic (Mini Contexts) Competence Test

no answer

there is no response

sociolinguistically inappropriate

the answer is not appropriate in the assigned context

sociolinguistically appropriate

the answer is appropriate in the assigned context

very good

the answer is appropriate in the context and also very good

N. B.

The detailed explanation of each label is exactly the same as the one in the conversational response test except that “sociolinguistic” has replaced “conversational”.

1. Speech Test

	below average	average	above average	very good
pronunciation				
grammar				
vocabulary				
content				
fluency				
discourse (logicality)				

2. Visual-Material Description Test

	below average	average	above average	very good
pronunciation				
grammar				
vocabulary				
content				
fluency				
discourse (logicality)				

3. Conversational Response Test

	no answer	conversationally inappropriate	conversationally appropriate	very good
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

4. Sociolinguistic (Mini Contexts) Competence Test

	no answer	sociolinguistically inappropriate	sociolinguistically appropriate	very good
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				