『教育研究』 39 (1997) pp.141–167
Educational Studies, 39 (1997) pp.141–167
International Christian University

Development of an English Speaking Test—— Establishing Validity through Multitrait–Multimethod (MTMM) Analysis

Yuji Nakamura

Theoretical background and rationale

With the advent of the Communicative Approach, the role of speaking ability has become more central in language teaching. Consequently, performance testing, especially testing speaking ability, has become one of the most important issues in language testing.

The increasing demand for the assessment of students' English speaking skills in Japan has forced teachers to recognize certain limitations in our understanding; for example, the construct of speaking. Since we are not sure exactly what the components of speaking are, it is sometimes difficult to explain what test scores mean, especially in relative terms.

Some scholars (Boldt and Oltman 1993; Henning and Cascallar 1992; Fulcher 1994) have approached the construct of oral proficiency, and explored the structure to some extent. However, there are still many limitations in this area because of the nature of speaking ability and the complexity of measuring spoken utterances.

Purpose of the research

With the goal of making a construct valid test of speaking ability in mind, the purpose of this research is to examine the following points through the technique of the multitrait–multimethod analysis:

1) Whether the proposed nine traits (pronunciation, grammar, discourse, fluency,content, vocabulary, comprehensibility, interactional competence, and sociolinguistic competence) are relevant and separable parts of speaking ability.
2) Whether there is a relationship between test methods and language ability. For this purpose, three methods (a writing test, an interview test, and a tape–mediated speaking test) were designed.
3) Whether the traits are fundamental elements of ability.
4) How the traits (with the same name across the three methods e.g. grammar in writing, grammar in the interview, and grammar in the tape test) are related.

The research will also attempt to discover the extent to which the proposed construct of speaking is reflected in other standardized tests such as the Test of Spoken English (TSE) or the American Council on the Teaching of Foreign Languages Oral Proficiency Interview (ACTFL OPI).

Research design and methods

The author made a matrix of three methods (writing test, interview test, and semi–direct speaking test) and nine traits (pronunciation, grammar, fluency, discourse, content, vocabulary, interactional competence, sociolinguistic competence, and comprehensibility) (See Appendix for the test). Thirty–three col-

lege students took the set of tasks(tests), and five English teachers, who are all native speakers of English, scored the test results using a 1–4 point scale rating sheet. The writing test results, the interview test results and the tape–mediated test results were rated by a group of two teachers each.

Factor analysis and cluster analysis were run to examine the construct of speaking ability. Correlations among the traits and the methods were used to investigate the convergence and the discriminance of the individual traits and methods. Discriminant validity was examined from two perspectives —different traits by same method and different traits by different methods. Convergent validity was examined from two perspectives——same trait by same method and same trait by different methods.

Establishment of Research Instrument Procedure

1) The operational definition of each trait is given below.

The general outline of the operational definition is whether a speaker can effectively express him/herself by using the following eight sub–traits.

* ——means examples of each trait.

(1) pronunciation

—— segmental features (individual sounds; vowels and consonants)

—— supra–segmental features (stress, rhythm, intonation)

—— volume (audible or not)

(2) grammar

—— tense and aspect

—— noun–verb agreement

—— word order

—— noun–personal pronoun agreement

(3) vocabulary

—— variety of words

—— word choice

—— idioms

—— nuances

(4) content

—— creativity

—— imaginativeness

(5)– (a) fluency in speaking

—— ease of speaking

—— speed

—— length, place and frequency of pauses

(5)– (b) fluency in writing —— intelligibility

—— how easy the composition is to read

—— how smoothly raters can read through the sentences

(6) discourse

—— cohesiveness

—— logical combination of sentences

(7) interactional competence

—— ability to use conversationally formulaic expressions

(8) sociolinguistic competence

—— ability to express his or her ideas in a sociolinguistically appropriate manner

(9) comprehensibility

—— the degree of intelligibility to the raters or overall impression

2) Making a matrix of traits and methods

*—— means cannot be measured.

** method in this research means task (a writing test, an interview, a tape–

mediated speaking test).

Method

	Writing test	/ Interview test	/ Tape– mediated speaking test
Trait			
1) pronunciation	——		
2) grammar			
3) fluency			
4) discourse			
5) content			
6) vocabulary			
7) interactional competence			
8) sociolinguistic competence			
9) comprehensibility			

Making each test

1) Writing Test

Students will be given a cartoon or a comic strip and given 15 minutes to write a composition of 100 words about the cartoon. There are two types of cartoons. One is a narrative type for evaluating students' grammar, fluency, discourse, content, vocabulary. The other is a conversation type for evaluating students' interactional competence and sociolinguistic competence. (See Appendix)

2) Interview Test

This is a 5-minute conversation with the classroom teacher following the suggested procedure below. The teacher will choose some of these questions which are relevant to probe each student's speaking ability.

1) Greeting

Self-introduction(Please introduce yourself by giving your name, the name of your high school and your hobbies.)

2) Short answer interview.

Where do you come from?

Do you have any club activities?

Do you have any brothers or sisters?

How do you come to school?

How long does it take to come to school?

Do you do a part-time job?

3) Long answer interview

Can you tell me more about your high school?

How many students are there?

Where is it located?

Can you tell me more about your family?

What do your family members do?

Can you tell me more about your club activities?

How many members are there?

How often do you meet with them?

Can you tell me more about your hobbies?

Why do you like them?

When did you start?

Tell me more about your part-time job.

What is your job?

What do you actually do?

What is your favorite TV program?

What kind of program is it?

Why do you like it?

What is the most interesting news these days?

Why do you think it is interesting?

What do you think about it?

4) Free form

Tell me more about your life at college

Do you talk to teachers? —— What are the teachers like?

Do you ever eat lunch at the dining hall? —— What is the dining room like?

Do you go to the library? —— What is the library like?

5) Winding up

Note:

We needed something to elicit students' sociolinguistic competence and interactional competence, such as a role play. However, it would interrupt the natural flow of conversation between the interviewer and student. Interviewers should include natural/ authentic questions in order to measure students' interactional and sociolinguistic competence in the greeting section and the winding up section so that interviewers do not need to stop the natural flow of conversation to start a role play abruptly. It is possible to say that the teacher– student conversation is one type of sociolinguistic context. Students' sociolinguistic ability can be measured through their ability to use formal expressions or polite expressions in the interview context. It may also be true that interactional ability can be rated through students' ability to handle necessary formulaic expressions in their responses to their interviewers.

Conclusion for the note.

Any test is a sample of the whole language proficiency. The interview test

is a sample of the whole language proficiency although it restricts the great possibilities of the measurement. Interview tests can be considered as samples of the whole sociolinguistic and interactional context in the real world. Therefore, the interview test is valid as a test tool to measure students' language proficiency.

We can think of the questions and responses that appear in the interview session as samples of sociolinguistic and interactional ability in the restricted context of a teacher–student relationship.

The interaction involved in the student–teacher interview allows their responses to be measured in terms of interactional ability. If students can interact with interviewers using formulaic expressions and polite expressions appropriately, then students can also have sociolinguistic / interactional competence.

3) Tape–mediated speaking test (See Appendix)

—— speech making test

—— interactional competence test

—— sociolinguistic competence test

Administration of the Test

1) Subjects

Thirty–three students of Tokyo Keizai University

2) Test administration

Tests were conducted in class.

Rating procedure and rating scale

1) A writing test was conducted by the researcher and evaluated by two raters (native speakers of English).

Rating Scale for Writing

(—— means cannot be measured)

	1	2	3	4
	poor			good
1) grammar				
2) discourse				
3) content				
4) vocabulary				
5) *fluency				
6) interactional competence				
7) sociolinguistic competence				
8) pronunciation	——			
9) overall impression (comprehensibility)				

* fluency in writing means: how easy it is to read, and how smoothly raters can read through it.

2) Interview tests were administered by two classroom teachers (native speakers of English) and evaluated by themselves.

Rating Scale for Interview

	1	2	3	4
	poor			good
1) pronunciation				
2) grammar				
3) discourse				
4) content				

5) vocabulary
6) fluency
7) interactional competence
8) sociolinguistic competence
9) overall impression (comprehensibility)

3) Tape-mediated tests were administered by the researcher and evaluated by two raters (native speakers of English).

Rating Scale for Speaking Test

	1	2	3	4
	poor			good

1) pronunciation
2) grammar
3) discourse
4) content
5) vocabulary
6) fluency
7) overall impression (comprehensibility)
8) interactional competence test
9) sociolinguistic competence test

Data Analysis

Factor analysis and cluster analysis were adopted 1) to examine the categorization of traits, 2) to investigate the relationship between the method

and the factor, 3) to explore the degree of contribution of the trait in each factor. Inter-trait correlation coefficients within each method were calculated in order to check the reciprocity of each trait internally within each factor. Also the correlations of the traits over three methods were examined to show the relationship among the traits.

Correlations among the traits and the methods were conducted to investigate the convergence and the discriminance of the individual traits and methods. Discriminant validity was examined from two perspectives — different traits by same method and different traits by different methods. Convergent validity was examined from two perspectives — same trait by same method and same trait by different methods.

Results and Discussion

1. Results of Inter-rater Reliability

The inter-rater reliability in each method was as follows:

Rater A vs. Rater B in Interview : .93

Rater C vs. Rater D in Tape Test: .77

Rater D vs. Rater E in Writing Test: .86

2. Results of Factor Analysis (Principal Component Factor Analysis)

Table 1 shows that three factors were extracted through the factor analysis, and these three factors agree with the different methods designed previously for this research as follows:

Table 1 Results of Factor Analysis

Note:Abbreviations are as follows:
Methods —— I: interview test, T: tape test, W: writing test
Traits ——— All: comprehensibility, Cont:content, Dis: discourse, Flue: fluency, Gra: grammar, Incom: interactional competence, Pron: pronunciation, Socom: sociolinguistic competence

	Factor 1	Factor 2	Factor 3
I_ALL	.93807	.04837	.17455
I_CONT	.72045	.16152	.39950
I_DIS	.90654	.10953	.23398
I_FIU	.84378	.20972	.22943
I_GRA	.76120	.16728	.09334
I_INCOM	.74474	.15113	.38222
I_PRON	.73095	.03046	.39340
I_SOCOM	.87389	.12819	.17067
I_VOC	.73233	.06267	.39272
T_ALL	.21794	.12093	.86280
T_CONT	.31485	.12894	.82531
T_DIS	.22268	.11845	.87409
T_FIU	.32271	.17259	.79521
T_GRA	.16685	.24088	.82018
T_INCOM	.36917	.49221	.33502
T_PRON	.38565	.24616	.65663
T_SOCOM	.41704	.38407	.51570
T_VOC	.33258	.17001	.81951
W_ALL	.18299	.91777	.03184
W_CON	.28561	.78377	.27079
W_DIS	.09690	.85977	.26881
W_FLU	.02865	.86566	.26218
W_GRA	.01422	.79431	.12939
W_INCOM	.02145	.85609	.22847
W_SOCOM	.11294	.78781	−.10288
W_VOC	.17243	.81338	.13860

Factor	Eigenvalue	Pct of Var	Cum Pct
1	12.48404	48.0	48.0
2	4.47094	17.2	65.2
3	2.45247	9.4	74.6

1) Factor 1 (Direct Oral Communication Ability) (DOCA)

The elements of this factor are comprehensibility, discourse, fluency, sociolinguistic competence, grammar, vocabulary, interactional competence,

pronunciation and content. We will name this factor Direct Oral Communication Ability. This ability deals with students' speaking ability in the face to face context, or students' reciprocal ability in the live context. The students utilize these linguistic elements for communication with some strategies such as eye contact or facial expressions. Therefore, students work on non-verbal signs as well as verbal messages.

2) Factor 2 (Written Communication Ability) (WCA)

The elements of this factor are comprehensibility, discourse, content, grammar, interactional competence, fluency, vocabulary and sociolinguistic competence. We call Factor 2 Written Communication Ability because students could communicate with each other through the written form. Students should take into consideration the context or the situation in which they are supposed to express their ideas.

Note: Although the factor loading of T-incom in Factor 3 is not as high as the ones in Factors 1 and 2, all of them are below .50. Therefore, there seems to be no significant difference among them.

3) Factor 3 (Semi-Direct Oral Communication Ability) (SDOCA)

The elements of this factor are discourse, content, comprehensibility, vocabulary, grammar, fluency, pronunciation, sociolinguistic competence and interactional competence. Factor 3 is named Semi-Direct Oral Communication Ability. This ability concerns students' speaking ability without being influenced by the external or the non-verbal elements. Students can/should focus on the monologue type communication by expressing their own ideas, or they can perform in English by concentrating on the in-coming information through their ears.

Writing is quite different from the interview and tape testing in the sense of a test mode. An interview and a tape test require audio dealing ability while writing does not. An interview and a tape test are distinct be-

cause some students feel nervous about speaking to a machine just mechanically, while other students do not feel easy when they must talk to an interlocutor in the interview session even with their teachers. Thus, the test practice can affect the difference of the two abilities.

The reason I have used the term "communication" for these three abilities is that students could communicate with each other in any of the three ways, by using verbal messages, by adopting non-verbal messages or by taking the written form. All are related to communication.

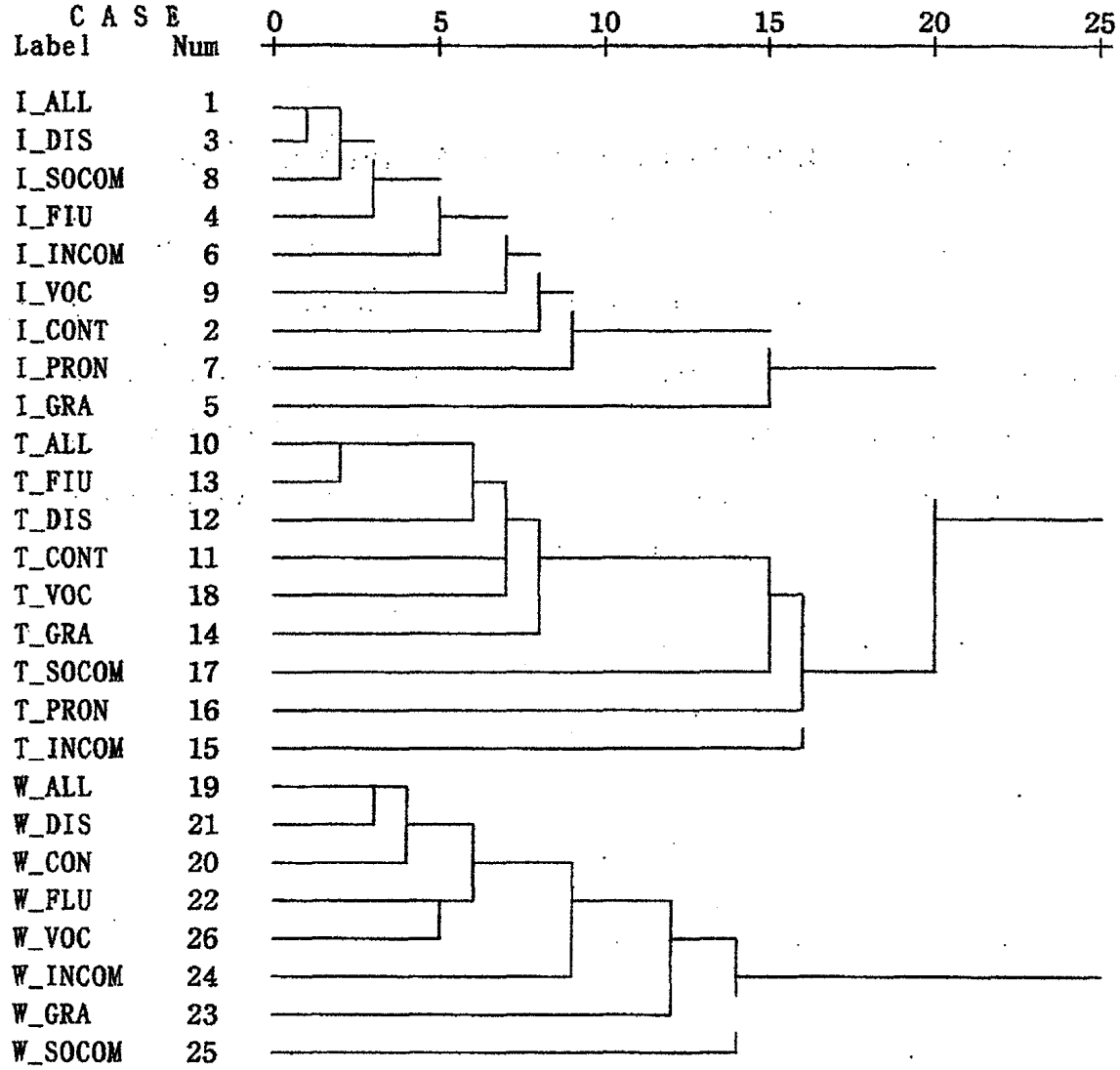
In the table, if there is a communication ability per se, the cumulative percentage of the variance suggests that 74.6% of the whole language ability ("communication ability") can be explained by these three factors. The remaining 25.4%, which cannot be clarified by the present results, may involve reading ability, listening ability (however, part of the listening ability is included in speaking ability) as well as strategic skills or cultural awareness which could be composite of the whole language ability.

3. Results of Cluster Analysis

Table 2 Result of Cluster Analysis

Note:Abbreviations are as follows:
Methods —— I: interview test, T: tape test, W: writing test
Traits ——— All: comprehensibility, Cont:content, Dis: discourse, Flue: fluency, Gra: grammar, Incom: interactional competence, Pron: pronunciation, Socom: sociolinguistic competence

Dendrogram using Single Linkage



The results of cluster analysis support the three factors obtained by factor analysis by adding more detailed information to the degree of the connection of each trait in each cluster. One of the findings is that three eventual clusters agree with the three factors. Another is that comprehensibility (overall impression) in each cluster is closely related with discourse and fluency. Whether the ability is speaking or writing, discourse and fluency are essential parts for the raters' comprehensibility.

Investigation of Individual Traits

4. Inter-Trait Correlations within the Same Method

Since there were no irregular traits (except T−incom in Factor 3) in the factor analysis for constructing the factors, each trait in each factor , as a whole, was contributing to form the reasonable factor. Also, they (the nine traits in each factor) were relevant as components of each factor. The question whether the traits were separable or not could be clarified through the following inter−trait correlations in Table 3(Table 3−1 to Table 3−3).

Table 3 Inter-Trait Correlations

Table 3-1 Inter-trait Correlations in Writing

Writing

	all	cont	dis	flue	gram	incom	socom	voc
all								
cont	0.83							
dis	0.85	0.83						
flue	0.78	0.76	0.81					
gram	0.72	0.61	0.71	0.67				
incom	0.77	0.72	0.74	0.74	0.65			
socom	0.68	0.46	0.51	0.59	0.63	0.71		
voc	0.75	0.70	0.74	0.82	0.60	0.69	0.56	

Table 3-2 Inter-trait Correlations in Interview

Interview

	all	cont	dis	flue	gram	incom	pron	socom	voc
all									
cont	0.71								
dis	0.88	0.78							
flue	0.81	0.68	0.84						
gram	0.68	0.56	0.64	0.69					
incom	0.76	0.78	0.82	0.80	0.56				
pron	0.75	0.72	0.73	0.56	0.61	0.65			
socom	0.86	0.68	0.82	0.75	0.66	0.64	0.77		
voc	0.79	0.63	0.76	0.71	0.54	0.63	0.62	0.68	

Table 3-3 Inter-trait Correlations in Tape Test

Tape Test

	all	cont	dis	flue	gram	incom	prom	socom	voc
all									
cont	0.78								
dis	0.80	0.70							
flue	0.86	0.77	0.77						
gram	0.67	0.71	0.78	0.66					
incom	0.33	0.50	0.44	0.39	0.48				
prom	0.62	0.68	0.68	0.60	0.63	0.60			
socom	0.57	0.58	0.57	0.51	0.61	0.68	0.57		
voc	0.76	0.80	0.80	0.72	0.77	0.40	0.67	0.69	

In this table, as the inter-trait correlations in writing(Table 3-1: the range is from .46 to .85), those in an interview test (Table 3-2:the range is from .54 to.88) and those in a tape test (Table 3-3:the range is from .33 to .86) show, the traits are all to some degree mutually interrelated and to some degree mutually exclusive. Therefore, it might be possible to say that they are, in part, separable.

5. Analysis of Same Traits in Different Methods and Different Traits in Different Methods

An investigation of the same-trait different-method analysis and the different-trait different-method analysis shows the following results. (See Table 4). The results show that correlations of the same traits across the three methods are not as high as had been expected.

Table 4 Correlations of the same traits across the three methods

Table 4-1 Correlations of the same and different traits between Writing and Interview

Writing vs. Interview

	Iall	Icont	Idis	Iflue	Igra	Iincom	Ipro	Isocom	Ivoc
Wall	0.26	0.29	0.28	0.33	0.21	0.29	0.19	0.30	0.22
Wcont	0.42	0.35	0.41	0.48	0.25	0.45	0.33	0.45	0.40
Wdis	0.21	0.34	0.25	0.30	0.15	0.35	0.21	0.29	0.19
Wflue	0.08	0.30	0.21	0.23	0.16	0.25	0.20	0.19	0.21
Wgra	0.08	0.24	0.11	0.14	0.27	0.23	0.16	0.11	0.10
Wincom	0.09	0.19	0.15	0.30	0.21	0.27	0.10	0.18	0.20
Wsocom	0.08	0.14	0.17	0.23	0.34	0.12	0.10	0.11	0.08
Wvoc	0.23	0.42	0.29	0.35	0.25	0.24	0.19	0.30	0.26

Table 4-2 Correlations of the same and different traits between Writing and Tape-Test

Tape vs. Writing

	Wall	Wcont	Wdis	Wflue	Wgra	Wincom	Wsocom	Wvoc
Tall	0.20	0.38	0.35	0.36	0.22	0.27	0.05	0.32
Tcont	0.21	0.44	0.40	0.37	0.17	0.35	0.05	0.24
Tdis	0.17	0.37	0.28	0.33	0.26	0.30	0.10	0.27
Tflue	0.27	0.48	0.41	0.38	0.24	0.31	−0.02	0.45
Tgra	0.22	0.42	0.39	0.39	0.36	0.46	0.19	0.27
Tincom	0.48	0.55	0.54	0.51	0.27	0.46	0.45	0.42
Tpro	0.33	0.52	0.51	0.38	0.30	0.29	0.14	0.25
Tsocom	0.42	0.45	0.42	0.43	0.27	0.47	0.45	0.40
Tvoc	0.25	0.42	0.39	0.30	0.31	0.37	0.14	0.27

Table 4-3 Correlations of the same and different traits between Interview and Tape Test

Tape vs. Interview

	Iall	Icont	Idis	Iflue	Igra	Iincom	Ipro	Isocom	Ivoc
Tall	0.37	0.50	0.41	0.40	0.35	0.48	0.53	0.34	0.48
Tcont	0.47	0.48	0.50	0.47	0.29	0.52	0.58	0.49	0.62
Tdis	0.36	0.54	0.41	0.45	0.26	0.56	0.49	0.32	0.55
Tflue	0.48	0.58	0.49	0.48	0.38	0.50	0.55	0.47	0.58
Tgra	0.24	0.56	0.40	0.41	0.30	0.55	0.46	0.26	0.40
Tincom	0.42	0.39	0.44	0.56	0.42	0.39	0.35	0.38	0.39
Tpro	0.53	0.49	0.49	0.48	0.37	0.52	0.59	0.54	0.53
Tsocom	0.42	0.57	0.57	0.63	0.50	0.49	0.37	0.45	0.51
Tvoc	0.48	0.61	0.50	0.49	0.36	0.61	0.54	0.49	0.54

One possible reason for this fact is that traits in each method are internally influenced by the nature of the factor and eventually presented the low correlation against the other traits in the other methods even though they have the same naming. Another possible reason is that the original definition of each trait might not be clear. Still another possible reason is that the trait in each method was basically different. In other words, there is a great possibility that grammatical ability in spoken form is different from grammatical ability in written form.

Comparison among an Interview Test, a Tape Test, ACTFL and TSE

Let us take a look at the components of the four tests before we compare them.

Components of an Interview Test

pronunciation, grammar, vocabulary, fluency, discourse, content, comprehensibility, interactional–competence, sociolinguistic–competence

Components of a Tape Test

pronunciation, grammar, vocabulary, fluency, discourse, content, comprehensibility, interactional–competence, sociolinguistic–ompetence

Components of ACTFL

pronunciation, grammar, vocabulary, fluency, pragmatic competence (= confident use of various conversation management devices), sociolinguistic competence (appropriate use of the major registers)

N.B. The idea of pragmatic competence is similar to that of interactional competence in the present research, and the idea of sociolinguistic competence resembles that of sociolinguistic competence in the present research.

Components of TSE

former version(before 1995)

pronunciation, grammar, fluency, overall comprehensibility

revised version(after 1995)

communicative language ability (communicative effectiveness)

—— comprehensibility

—— linguistic accuracy (pronunciation, grammar, fluency, vocabulary)

—— coherence and use of cohesive devices

—— appropriateness of response to audience/situation

—— addressing the communicative task

Judging from the results of this research, an interview test and a tape test are measuring the whole communication ability from different modes. They might cover some common linguistic areas while there are some distinct areas that the counter part cannot compensate for. This might be true for the situation between ACTFL and TSE. ACTFL through an interview test and TSE through a tape test seem to play a similar role as the interview test and the tape test do in the present research for measuring the whole communication ability.

Since, through this research, we were able to find the indispensable role

played by both the direct and the semi–direct speaking tests, it is not as difficult to presume that ACTFL and TSE will function in the same way because of the similar components of these speaking tests.

Conclusions

Three factors obtained from the factor analysis corresponds with three methods designed for measuring the ability. These three factors were named Direct Oral Communication Ability, Semi–Direct Oral Communication Ability and Written Communication Ability individually according to the nature of each ability.

As far as speaking ability is concerned, the two modes mentioned above are rather distinct judging from the factor analysis. However, both are measuring a common area of communication ability to some extent.

Nine traits in Direct Oral Communication Ability (Interview test) contribute to become the fundamental elements of the factor while they still maintain their own characteristics which cannot be replaced by other traits. This idiosyncrasy was made clear through the inter–trait correlation coefficients. The same is true for the nine traits in Semi–Direct Oral Communication Ability. The nine traits function as factor construct elements but they still have their distinctiveness.

The traits in each factor, even though they have the same naming (grammar, vocabulary, discourse etc.), are working separately in the three methods. One possible reason for this phenomenon is that the trait, influenced by the other traits internally within the factor, produced a new phase of the trait. Another possible reason is that the trait in one factor is basically different from the other two traits in the other factors, even though they have the same name. In other

words, it would be possible to coin terms such as"grammar in face to face situations,"and"grammar in the tape test situation." The solution to this problem is left to the results of future study.

One way to look at the distinctions of the methods can be explained as follows: The lower or intermediate–lower level students do not necessarily demonstrate their competence in all language abilities. Some are poor at writing, some are good at interview tests and others are skillful in speaking without an interlocutor. It is said that unlike these lower or intermediate lower level students, advanced or high intermediate level students show their constantly high language competence in various language skills. This fact might be one of the causes of the distinctions of the three factors with the subject in the study.

Interview tests could measure part of the whole communication ability, and tape tests could measure another part of the whole communication ability. The traits in each method can be a good composite of the ability the method is measuring.

One interesting finding is that tape tests and interview tests have individual purposes and neither of which can be replaced by the other. Therefore, in addition to the distinction between the multiple choice test and performance test, the difference between a tape test and an interview test even in the performance test should be noted, and each test should be used to play a complementary role in other tests.

We were able to find some similarities between the present interview test and ACTFL interview test, and between the present tape mediated test and TSE. Further research into concurrent validity will hopefully enforce the relationship among these tests.

Finally, the following points should be taken into account for future study.
1) The number of testees should be increased to get more generalizable results.
2) A much higher inter–rater reliability within each method should be obtained.

3)More exhaustive study of traits and methods should be undertaken.

Note: A version of this paper was presented at the 30th TESOL Annual Convention, Chicago, March 26–30, 1996 and at the 11th World Congress of Applied Linguistics, Jyvaskyla Finland, August 4–9, 1996. This research was funded by the Ministry of Education (A Research Grant of the Ministry of Education: Grant No. 07610487).

Bibliography

ACTFL (American Council on Teaching of Foreign Languages). (1989). *ACTFL testers training manual.* New York: American aouncil on Teaching of Foreign Languages.

Bachman, L. F. (1988). Problems in examining the validity of the ACTFL Oral Proficiency Interview. *Studies in Second Language aquisition,* 10, 149–163.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F. and Savignon, S. J. (1986). The evaluation of communicative language proficiency: a critique of the ACTFL Oral Interview. *Modern Language Journal* 70,4, 380–390.

Baker, D. (1989). *Language Testing: A Critical Survey and Practical Guide.* Edward Arnold.

Boldt, R. F. (1992). Reliability of the Test of Spoken English revisited (*TOEFL Research Report* No. 40). Princeton, NJ: Educational Testing Service.

Boldt, R. F., Larsen–Freeman, D., Reed, M.S., & Courtney, R. G. (1992). Distributions of ACTFL ratings by TOEFL score ranges (*TOEFL Research Report* No. 41). Princeton, NJ: Educational Testing Service.

Boldt, R. F. & Oltman, P. K. (1993). Multimethod construct validation of the

Test of Spoken English (*TOEFL Research Report* No. 46). Princeton, NJ: Educational Testing Service.

Brown, J. D. (1996). *Testing in language programs*. New Jersey: Prentice Hall Regents.

Cohen, A. (1994). *Assessing language ability in the classroom*. Second edition. Boston Mass.: Heinle and Heinle.

Dandonoli, P. and Henning. G. (1990). An Investigation of the Construct Validity of the ACTFL Proficiency Guidelines and Oral Interview Procedure. *Foreign Language Annals* 23,1,11–22.

ETS (Educational Testing Service). (1982). *ETS oral proficiency testing manual.* Princeton NJ: Educational Testing Service.

ETS (Educational Testing Service). (1990). *TSE manual for sore users.* Princeton NJ: Educational Testing Service.

Fulcher, G. (1994). Some Priority Areas for Research in Oral Language Testing. *Language Testing Update* 15, 39–47. International Language Testing Association.

Henning, G. (1987). *A guide to language testing*. Cambridge, MA: Newbury House.

Henning, G. (1992). The ACTFL Oral Proficiency Interview: Validity Evidence. *System* 20,3,365–372.

Henning, G. & Cascallar, E. (1992). A preliminary study of the nature of communicative competence (*TOEFL Research Report* No. 36).Princeton, NJ: Educational Testing Service.

Henning, G., Schedl, M., & Suomi, B. K. (1995). Analysis of proposed revisions of the Test of Spoken English (*TOEFL Research Report* No. 48). Princeton, NJ: Educational Testing Service.

Pienemann,M., Johnston,M., and Brindley, G. (1988). Constructing an Acquisition–Based Procedure for Second Language Assessment. *Studies in Sec-*

ond Language Acquisition 10, 217– 234.

Appendix A (for Writing)

1) Please write a story in English based on the picture narrative.

2) Please write a conversation in English based on the picture narrative.

Appendix B (for a Tape Test)

1) Speech Test

Please choose one topic you want to talk about among the eight topics given below and give a two–minute speech about it. (my friends, my family, part–time work, my hobbies, traveling fashion telephone conversations, college life)

2) Interactional Competence Test

Please give a quick and appropriate response in English to each sentence or question which will be heard through your headphones.

—— Nice to meet you.

—— What is your name?

—— Could you spell it please?

—— How are you?

—— What do you do?

—— Can you tell me the time?

—— What is the date today?

—— What is the weather like today?

—— What do you usually do on Sundays?

—— How do you come to school?

3) Sociolinguistic Competence Test

Please give an appropriate response in English in each context.

—— Context 1(Apologizing and making an excuse)

—— Context 2(Complaining and requesting)

—— Context 3(Asking for repetition)

—— Context 4(Interrupting)

—— Context 5(Telephoning: taking a message)