DOCUMENT RESUME

ED 417 598 FL 025 120

AUTHOR	Nakamura, Yuji
TITLE	A Study of Raters' Scoring Tendency of Speaking Ability through Verbal Report Methods and Questionnaire Analysis.
PUB DATE	1996-09-25
NOTE	17p.
AVAILABLE FROM	The "Journal of Communication Studies" is published by Keizia University, Tokyo, Japan.
PUB TYPE	Journal Articles (080) -- Reports - Research (143)
JOURNAL CIT	Journal of Communication Studies; n5 p3-17 Sep 1996
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*English (Second Language); *Evaluation Criteria; Foreign Countries; Interrater Reliability; Language Research; *Language Tests; *Rating Scales; *Scoring; Second Language Instruction; Speech Skills; Student Evaluation; Surveys; Testing; *Verbal Tests
IDENTIFIERS	Japan

ABSTRACT

To find ways to improve rater reliability of a tape-mediated speaking test for Japanese university students of English as a Second Language, two studies gathered information on: how raters actually made their choices on rating sheets of students' speaking ability; determined what criteria teachers think they use and actually use in rating students' speaking ability; and drew implications for improving rater reliability. In the first study, subjects were six native English speakers and three Japanese teachers of English. They answered a 16-item questionnaire about rating techniques and criteria, then four months later, rated 30 student tapes, answered the questionnaire again, and gave a retrospective self-report about the internal process of evaluating student skills. One of the subjects also gave a self-report concurrent with rating tapes. In the second study, six audio- and videotape raters analyzed their rating tendencies through verbal reports, self introspective and retrospective reports, and interviews. Results of the two studies and implications for teaching and testing are reported. The instruments (questionnaires, retro/introspective questions, and interview questions) are appended. Contains 3 references. (MSE)

A Study of Raters' Scoring Tendency of Speaking Ability through Verbal Report Methods and Questionnaire Analysis

Yuji Nakamura

コミュニケーション科学第5号所載抜刷（1996年9月）

A Study of Raters' Scoring Tendency of Speaking Ability through Verbal Report Methods and Questionnaire Analysis

Yuji Nakamura

Theoretical background and rationale

Verbal report techniques have been used to investigate more directly the criteria that raters actually draw on when making their ratings (Cohen 1988; 1994). There has been research on the use of the verbal report methods (think aloud, introspective and retrospective self-observation, and self-report) from which to gather information on how raters actually make their choices. Some of the work on collecting verbal report data comes from raters in the area of L2 writing. It would also be helpful to have verbal report data to determine just what criteria teachers actually use, as well as those that they think they use.

Purpose of the research

To find ways to improve rater reliability of a tape mediated speaking test for Japanese university students, the technique of the verbal report by the raters and the questionnaire analysis were utilized: 1) to gather information on how raters actually made their choices on their rating sheet of students' speaking ability, 2) to determine what criteria teachers actually use, as well as those that they think they use in rating students' speaking ability, and 3) to make the best use of the results for the training of raters, for both the classroom teaching and classroom learning. Two research questions are:

1) Is there a difference in teachers' ideas about the degree of importance of the various components of speaking ability?
2) What is happening in the raters' mind or what do teachers value in the process of evaluation of students' speaking ability?

Research design and methods

1. Subjects

Nine (six native speakers of English and three Japanese teachers of English) were employed as subjects.

2. Material

A questionnaire (1-4 scale) consisting of 16 items (e.g. pronunciation, grammar, vocabulary) concerning the degree of importance of the sub-categories of speaking ability was used for the retrospective questionnaire analysis. Audio/video tapes, in which 30 students' answers were recorded, were used for the simultaneous introspective self-report and retrospective self-report analysis.

3. Procedure

1) Raters answer the questionnaire before rating the tapes.
2) Four months later, these same raters score the tapes.
3) While rating video tapes, one teacher will give an introspective self-report about the internal process of the evaluation in his mind.
4) Raters answer the same questionnaire again after finishing the tape evaluation.
5) Raters are asked to give a retrospective self-report about the internal process of their evaluation of the students' speaking ability.

Study 1 (Questionnaire Analysis)

Subjects

Nine teachers participated in the questionnaire research. Seven of them answered the questionnaire (see Appendix) before and after the audio tape evaluation, while the other two answered it before and after the interview session.

They answered the first questionnaire four months before the tape evaluation and they answered the questionnaire again just after the tape evaluation. It is hoped that the raters had forgotten their answers to the first questionnaire by the time of the tape evaluation. Also it is hoped that their answers to the second questionnaire were exactly the same as those which were in their mind during the process of the tape evaluation.

Material

A questionnaire consisting of 16 items was used for the pre and the post evaluation.

Results and Discussion

Table 1 shows the results of the t-test of the questionnaire responded by the rater between the pre and the post evaluation. The results of the t-test show that there is no statistical difference between the pre rating and the post rating. This means that the degree of importance of each item to the rater has not changed; in other words, there has been no change in the idea of the construct of speaking before and after the evaluation.

Since there is no difference between the two stages of the evaluation concerning the idea of the construct of speaking, the inter-rater reliability will increase more if we have the training session before the evaluation and come to a confirmed agreement. Accordingly, each rater will consistently evaluate tapes within the confirmed framework.

Table 1

The results of the t-test (by raters) of the questionnaire responded by the rater between the pre and the post evaluation

Rater	Pre		Post		t	p
	M	SD	M	SD		
A	2.75	1.07	2.94	0.77	−0.57	n.s.
B	3.06	0.77	2.50	1.21	1.57	n.s.
C	3.00	0.63	2.75	0.45	1.29	n.s.
D	3.25	0.78	3.13	0.72	0.47	n.s.
E	2.50	1.27	2.50	1.21	0	n.s.
F	3.06	0.68	3.00	0.52	0.29	n.s.
G	3.31	0.70	3.31	0.95	0	n.s.
H	2.81	0.83	2.81	0.75	0	n.s.

(p>.05) (df=30)

Furthermore, if in the training session, the problem of rating criteria as well as the problem of the construct becomes clear enough, the more reliable evaluation with the confirmed rating criteria and the confirmed construct of speaking will be obtained.

So far, we have conducted the training session of the rating criteria before the real tape evaluation, but we rarely go into the discussion of the construct of language proficiency. The results of the present research convince us of the importance of the confirmation not only of the rating criteria but also of the construct, because raters are not biased by the tape material concerning their idea of the construct of the language ability.

There was no statistical difference in each rater. One rater showed the identical response between the pre and the post scoring, which is quite unlikely to happen in the usual research. Therefore, this rater's data was eliminated from the following statistical analysis.

Since there is no statistical difference in the eight individual raters between the pre and the post stages, it seems that they claim they evaluate the tapes as they think they have to.

If we investigate the change of the mean score (4 people declined; 2 people increased; 2 people stayed in the same), there seems to be a declining tendency overall. If we check the change of the SD (5 people spread wider; 3 people shrink together), there seems to be a tendency toward spreading overall.

There was no difference between the ideal (the pre) and the real (the post) evaluation. If the questionnaire functions well in the training session, and the raters unconsciously keep the content of the questionnaire in mind, then they try to stick to the criteria of the evaluation. In other words, there must be a "washforward" effect.

Since there was no difference between the pre and the post evaluation , the concept of the pre evaluation continued in the real evaluation and also in the post evaluation. Therefore, it is possible to imagine that the continual concept from the pre evaluation was still working in this real evaluation and raters evaluated tapes in almost the same way as they had done in the ideal pre situation.

Table 2 demonstrates the results of the t-test of the questionnaire responded by the rater between the pre and the post evaluation. If we compare the mean score of the items between the pre and the post evaluation (9 items declined; 4 items increased; 3 stayed in the same), there seems to be a declining tendency. If we compare the SD of the items (5 items declined, 8 increased; 3 stayed in the same), there seems to be a tendency toward spreading.

Since there is no significant difference, the concept of the pre stage continues in the real evaluation, especially in the global evaluation of speaking ability. This unconsciously embedded concept shows up in the evaluation.

Each item was taken in the evaluation as equally as it was in the questionnaire. However, the item "pronunciation" had almost a significant difference at the .05 level, which means that in pronunciation, the real evaluation is a little different from the ideal

Table 2

The results of the t-test (by items) of the questionnaire results responded by the rater between the pre and the post evaluation

item	Pre		Post		t	p
	M	SD	M	SD		
1	3.33	0.71	2.78	0.44	2.00	n.s.
2	3.11	0.60	2.89	0.93	0.60	n.s.
3	3.11	0.93	3.11	0.33	0	n.s.
4	3.11	0.78	3.22	0.67	−0.32	n.s.
5	3.22	0.67	2.78	0.67	1.41	n.s.
6	3.00	1.00	2.78	1.20	0.43	n.s.
7	3.22	0.97	3.22	0.97	0	n.s.
8	3.22	1.09	3.44	0.53	−0.55	n.s.
9	3.56	0.53	3.44	0.73	0.37	n.s.
10	2.44	0.88	2.67	1.23	−0.44	n.s.
11	2.78	0.67	2.67	0.87	0.31	n.s.
12	1.78	0.44	1.89	0.78	−0.37	n.s.
13	2.22	0.67	2.00	0.50	0.80	n.s.
14	3.78	0.44	3.78	0.44	0	n.s.
15	3.67	0.50	3.56	0.53	0.46	n.s.
16	2.78	0.67	2.67	0.71	0.34	n.s.

(p>.05) (df=16)

construct.

When the questionnaire was given before the tape evaluation, the content of the construct was unconsciously embedded into the raters' mind and was retained through the tape evaluation, even after the tape evaluation. Therefore, if we take advantage of this phenomenon, and give a comprehensive training session where raters can come to an agreement on the construct of speaking for the global evaluation, the raters will be more reliable and the inter-rater reliability will be much higher.

Summary of Study 1

1. There was no statistical difference in individual raters and items between the pre and the post evaluation. However, as a whole, there seems to be a tendency toward declining.
2. Some items, such as pronunciation, showed a declining tendency of scoring.

3. Individual raters indicated the consistency of raters.
4. There was no statistical difference in the intra-rating.
5. There was no statistical difference in individual items.
6. There was no statistical difference in raters. One idiosyncratic person whose responses in two ratings were identical, was subsequently eliminated in the statistical analysis.
7. As there was no statistical difference, we could say that raters were scoring tapes in almost the same way that they answered the questionnaire.

Conclusion for Study 1

Since there is a consistent rating tendency in the intra-rater scoring, if there is an agreement among raters (inter-rater) through a comprehensive training session, (whether it is or group training or an individual training), we will be able to obtain much higher inter-rater reliability.

Even though it is very difficult to investigate an intra-rater reliability because of practical reasons, we could replace it with the verbal report analysis, using the questionnaire to get an overall idea.

Even though there is an assumption that an ideal evaluation and the real evaluation should be different, this research shows no statistical difference. This is probably because raters answered the pre-evaluation questionnaire by taking the practical limitations into consideration. In fact, there may not be a big difference between the two.

Study 2 (Analysis of Rating Tendency through Verbal Report, Self-Introspective, Self-Retrospective, Interview)

Data Source

1. Data came from two interviewer-raters through retrospective verbal reports. After the interview test was over, these two raters were asked to look back and report on what was happening in their rating during the interview.
2. Other data came from a video tape rater through a verbal report. In the process of the videotape evaluation, the rater was asked to speak out about what is happening in his mind as much as possible.
3. Other data came from three audio tape raters through retrospective verbal reports. After the tape evaluation was over, the raters were asked to recollect what was happening in the process of rating.

Results and Discussion

1) Analysis of the data from interviewer-raters

The interviewer was conscious about the greeting portion (how to start the interview, how to relax the student, how not to be monotonous as an interviwer). The interviewer changed the questions from time to time in content but not level, so that the interviewer would not get bored.

The content of the questions in the interview is highly dependent on how good the student teacher/interviewer relationship is before the interview. The big difference between the speaking test and the writing test is that the interviewer should have good rapport with the students and create as comfortable an atmosphere as he can before or within the interview session. There is a difference in terms of asking questions between the situation in which an interviewer knows the students well and the situation in which an interviewer knows them very little or not at all.

Even in the test situation, the grading is greatly influenced by the content of the students' utterances.

It is important to have students ask questions towards the end of the interview because this part can compensate for the lost element of the reciprocal communication. Communication should be reciprocal and the interview can be reciprocal communication if the questioning functions well when the students participate.

It seems that the interviewer (the rater) is unconsciously assessing the students' speaking ability by their listening ability. Although the rater gives high points to an appropriately quick response, the rater tends to give high points to the students whose utterances have rich or interesting content. Also, the rater gives higher points to the students who could expand the topic spontaneously.

The rater was not focusing on discrete/individual items in the interview. The decision making about the level assessment was done when there was a breakdown in communication (long pause, silence etc.). These were important factors for the evaluation. The main things that the interviewer was thinking about in the process of the interview are as follows:

1) How many minutes are left?
2) How can I get this student to speak up?
3) How can I elicit the student's speaking ability?

In other words, there were many cases when the interviewer decided the topic or decides to change the level and the content of the topic, indicating that the interviewer has already finished the assessment of the students' speaking ability.

One aspect of speaking ability in the interview session is what kind of question the student can understand and speak about. There are some other examples such as 1) Can the student follow the speed of the question? 2) Can the student follow the topic or the content? 3) Can the student understand the content of the question and express his/her opinions on it?

The interviewer was thinking about or preparing the next question while the student was responding. (At that moment, the continuous part of the question was at the same level, and the next question was being produced in the interviewer's mind.)

The big difference between ACTFLOPI (the American Council on the Teaching of Foreign Languages Oral Proficiency Interview) and this type of interview is that, in ACTFLOPI, the criterion is previously decided on and the interviewer-rater does not go into the content of the item (just level checking and probing). In other words, the interviewer is more interested in the level assessment. However, in this type of class interview, the interviewer tries to get the students to speak up. Therefore, the interviewer-rater always thinks about what relevant question will elicit the students' speaking ability and thus, the interviewer is deeply involved in the content of the students' production as well. This is where the instability of the question-preparation (the inconsistency of the question-presenting), occurs. The reliability becomes even lower when different questions are asked to different students. There exists a dilemma between trying to have students speak up and not being able to keep reasonable reliability.

The interviewer cannot check discrete/individual items in the interview session. They have to do it later or have another rater do the grading.

There is a difference between a situation when an interviewer is a known classroom teacher and a situation when an interviewer is a complete stranger. In a classroom setting, classroom teachers interview the students and grade their speaking ability. In other words, we have to think about the context of the questions and also the check sheet. Also, the test date (at the beginning of the year, or in the middle of the year, towards the end of the year) should be taken into account because the greeting words or question words would be completely different depending of how well the interviewer knows the students, as mentioned above.

If the holistic evaluation in the interview session is the same as the in-class evaluation, the interview should be used for other purposes such as discrete point evaluation. For that purpose, the interview session should be recorded or video-taped and the additional items must be included in the check sheet.

In the interview, if we change the questions depending on the level of the students, then we might not be able to make a fair judgment. How can we compare the two students who got 3 points by being asked quite different questions? When the interviewer changes the question or the level of the topic, the interviewer already changes the criteria. It seems that the interviewer's next question is suggesting the students' listening/speaking/interactional level. If the question is very difficult or at natural speed, then it may be beyond the student's ability, and if the question is paraphrased or made simpler, this indicates that the students' listening ability (speaking ability is included in this aspect) is rather low. Speaking ability seems to be assessed by the listening ability in the interview session. The content of the interviewer's questions indicate the interviewee's level of speaking ability.

Some raters claim that interactional level formulaic expressions can easily be mastered by being exposed to native speakers in foreign countries, so these are not so highly valued. However, the ability to make oneself understood with content in the utterances is very important in speaking. In other words, comprehensibility to the raters from the interviewer's points of view is important.

2) Analysis of the data from a video-tape rater through an introspective verbal report

The rater tried to share the context or the situation in which the student was involved by trying to understand what the student was trying to say.

The rater carefully looked at students' eye contact, facial expressions because he needed to know if the student understood the question. Since some students do not know how to make a response verbally, when they do not understand the complicated question or unfamiliar words, facial expressions or hand movements can be a good indicator to show that they are stuck with the listening part. In other words, when the rater of the video tape looks at the facial expressions, body movements, and eye contact very carefully, he or she focuses on the students' listening ability rather than on the speaking ability. This is due to the way the students show their understanding of the question in their behavior, especially in their eyes.

The rater sometimes got impatient or frustrated with the students' low voice or vagueness in utterances. He was checking if the student was able to expand the topic and initiate the conversation rather than just waiting for and answering the questions. He checked if the pronunciation was clear, if the responses were appropriate, and if the utterances were natural.

Comprehensibility is very important. Students should try to make him/herself understood in English, while raters should try to understand the students. Students can improve by polishing pronunciation, grammar, vocabulary, discourse, sociolinguistic knowledge etc., whereas raters (teachers) can improve by expanding their knowledge of the topics which students are familiar with. Thus, if the two sides meet in the shared context, real communication occurs. In other words, both students and teachers should be listener friendly. Students should be listener friendly by showing their efforts to make themselves understood in English, while teachers should try to grasp the topics that students are interested in so that teachers can easily share the topic with the students, even in the test situation. Moreover, teachers should have as much information as possible about the students' academic background.

One aspect of measurement which the Mombusho stresses in the new course guidelines is the students' positive attitude toward communication. Although this is intended for the students' attitude towards communication, the raters/teachers should reconsider themselves by taking into account the content of comprehensibility. Students, as well, should try to speak as clearly as possible so that the workload of the raters will be diminished.

3) Analysis of the data from three audio tape raters through an retrospective verbal report

Raters tried to understand what the students wanted to say. It was not the content but rather the comprehensibility that mattered in the speaking test. The raters' focus on the evaluation was not on grammar or pronunciation, but on total communication, in other words, comprehensibility.

The important thing for the raters is how much background knowledge we can share with students. Raters need to try to increase their knowledge of topics in which students are interested in everyday classroom situations, even though it is sometimes possible to elicit the new information from the students in the test situation. We must take into consideration that the beginning level or intermediate low level students are more likely to speak out about their familiar topics. This can happen only when the raters elicit the topic and share it with the students.

One way is to relate the topic the students choose to what you know about it as much as possible. In other words, how much can raters use the script that the students choose.

In one sense, how to be a good rater is similar to how to be a good listener. The component of listening ability resembles the component of the rating ability.

Students should try to make themselves understood in English by improving their language skills such as, grammar, vocabulary, discourse, fluency etc. so that raters will want to listen to them more. They should try to show their effort at making themselves understood in English.

Raters, although they are sometimes intrigued by the students' mistakes in grammar and pronunciation in the process of evaluation, should keep consistently on the track of comprehensibility by trying to share the common ground related to the topic, chosen by the students.

Furthermore, since raters want to share the topic/context with the students through tape by trying to understand what the student is trying to say, they should also expand the topics in the testing situation as well as in the classroom situation so that students can understand them more.

Summary of Two Studies (Study 1 and Study 2)

Study 1 (The questionnaire results)

In terms of the raters, there was no statistically significant difference between the pre and the post tape evaluation concerning the degree of importance of each item to the raters in the questionnaire. In other words, there is no difference in the raters' mind between the ideal construct of speaking and the real/practical construct of speaking.

Although there was no statistical difference in raters' mind about the importance of the items between the pre and the post evaluation, there was a tendency in that the mean became lower and the SD less spread. After the tape evaluation, they became more realistic.

From the viewpoint of individual items, there was no significant difference between the two stages, though there was a tendency that the post evaluation rating was generally lower than the pre in each item.

Raters tended to value comprehensibility and stuedents' effort by "trying to speak only in English" or "trying to make themselves understood in English."

Since there was no difference between the pre and the post evaluation, there was no difference in the importance of each item between the classroom situation and the test situation. It is hoped that the teaching points in the classroom are reflected well in the testing points, which makes the content validity of the test much more important. Furthermore, the fact that what is stressed in the classroom is tested in the test will give a positive washback effect to the students. In other words, the agreement between the teaching points and the testing points will motivate the students to study more by aiming at the goal set by the teachers. This is because we, as human beings, are greatly affected by the content of the test.

However, there is one condition that needs interpretation in this agreement. Teachers/raters must always update the construct of speaking ability by taking into consideration the linguistic framework of speaking ability, students' needs, needs from the society etc. Otherwise the established test might lead the students to an unexpected goal.

The well established test along with the well organized training session which provides teachers with a common construct of speaking ability and a common rating scale will give a fair judgment and a positive washback effect to students.

Study 2 (The verbal report results)

For the interview test interviewers should ideally be different from raters, and classroom teachers should not be the raters of their own students.

Interview-raters, audio-tape raters, and video tape raters rate comprehensibility (to try to understand what the students try to say) highly. In order to make this comprehensibility possible, teachers should expand the topics or the information students are interested in so that they can share them even in the testing situation. Students should be taught to be listener friendly by paying more attention to the comprehensibility to the listener. This will naturally lead them to realize the importance of making themselves understood in English, and that the key to this is the improvement of basic language skills such grammar, vocabulary, pronunciation, fluency and discourse.

Overall Conclusion for Study 1 and Study 2

Raters should realize that they need to share with the students either the topic or the context so that the students can expand even in the test situation and in general, learn how to speak out more comfortably. Also, students (test takers) should be reminded that they can be listener friendly as well by trying to make themselves understood in English with a clear and audible tone, also they should be more open to subject area change. Cooperative attitudes between the two sides will make the communication easier and the judgment will be much more accurate.

In everyday classes, teachers should be more concerned with the comprehensibility/communication rather than the minor grammar mistakes. If raters or teachers are consistent in their policy of teaching and testing by focusing more on comprehensibility, then the content validity of the test will be much higher. Accordingly, there will be a better washback effect of the test, and the students will study more for communication by making an effort to make themselves understood in English. In other words, comprehensibility builds the communication.

Communication occurs when comprehensibility reaches an acceptable level. Comprehensibility to the listener increases when the students become listener friendly, and the listener can share the topic, information, and context with the students. Communication requires a cooperative attitude between the students and the teachers. Therefore, cooperation between teachers and students should occur not only in the teaching/learning context in the classroom but also in the test situation, so that the fair judgment of communication ability will be possible. What the result of the present research shows us is that it is not only students but also teachers that build up the testing context in which natural communication and fair judgment are expected and strived for.

Acknowledgment: This research was supported by the Hasegawa Language Education Promotion Fund.

References

Cohen, A. D. (1994). Assessing language ability in the classroom. Boston, Mass: Heinle & Heinle.

Cohen, A. D. (1988). The use of verbal report data for a better understanding of test taking processes. Australian Review of Applied Linguistics, 11(2), 30–42.

Weigle, S. A. (1994). Effects of training on raters of ESL compositions. Language Testing, 11(2), 197–223.

Appendix I

Questionnaire

When you assess Japanese students' speaking ability, how much weight do you put on each category below? Please circle one choice for each category. If you would like to add more items to this list, please feel free to do so in the blank space.

1: not important at all
2: not very important
3: rather important
4: very important

1)	pronunciation	1	2	3	4
2)	grammar	1	2	3	4
3)	discourse	1	2	3	4
4)	content	1	2	3	4
5)	vocabulary	1	2	3	4
6)	fluency	1	2	3	4
7)	interactional competence	1	2	3	4
8)	sociolinguistic competence	1	2	3	4
9)	comprehensibility	1	2	3	4
10)	ability to speak immediately	1	2	3	4
11)	ability to speak continuously	1	2	3	4
12)	ability to use gestures appropriately	1	2	3	4
13)	ability to use appropriate eye contact	1	2	3	4
14)	ability to make him/herself understood	1	2	3	4
15)	ability to try to speak only in English	1	2	3	4
16)	ability to speak with comfortable speed	1	2	3	4

Appendix II

1. Questions used for the interviewer-rater

1) What was happening in your mind when you were interviewing the first student? (at the beginning, in the middle and the towards the end of the interview)

2) What was happening in your mind when you were interviewing the second student?

3) Can you tell me who was the most interesting student or the most impressive student?

4) In the process of one interview (an interview for one person), what were you doing during each of the three stages (at the beginning, in the middle and the towards the end of the interview)

5) When you were asking questions or listening to students' responses what were you thinking about?

6) What was the big difference between what you anticipated and what you were actually doing in the interview session, in terms of evaluation items or evaluation criteria?

7) When you changed your questions, or the level of difficulty, did you change your criteria (e.g. from 3 to 2)?

8) Was it difficult to be an interlocutor and at the same time a rater?

9) What were you actually thinking about when the student was making a response?

10) Was there any unexpected finding in/after the interview ?

2. Questions used for the tape(audio/video) raters.

1) What was happening in your mind when you were evaluating the first student?

2) What was happening in your mind when you were evaluating the second student?

3) In the process of evaluation, what steps were you going through?

4) What was the big difference between what you had anticipated before and what happened after the tape evaluation?

5) What was the main problem or the most difficult thing as a rater?

6) Was there any unexpected finding in/after the tape evaluation?

Kathleen M. Marcos, Acquisitions Coordinator
ERIC Clearinghouse on Languages and Linguistics
1118 22nd Street, NW
Washington, DC 20037
Tel: 800-276-9834 / 202-429-9292
Fax: 202-659-5641
E-Mail: kathleen@cal.org

****

Visit our web site at http://www.cal.org/ericcll
REPRODUCTION RELEASE FORM FOLLOWS:

ERIC REPRODUCTION RELEASE

I. Document Identification:

Title: A Study of Raters' Scoring Tendency of Speaking Ability through Verbal Report Methods and Questionnaire Analysis

Author: Yuji Nakamura

Corporate Source: Tokyo Keizai University

Publication Date: September 25, 1996

II. Reproduction Release: (check one)

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in Resources in Education (RIE) are usually made available to users in microfiche, reproduced in paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. If permission is granted to reproduce the identified document, please check one of the following options and sign the release form.

[✓ Level 1 - Permitting microfiche, paper copy, electronic, and optical media reproduction.]

Level 2 - Permitting reproduction in other than paper copy.

Sign Here: "I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit

reproduction by
libraries and other service agencies to satisfy information
needs
of
educators in response to discrete inquiries."

Signature: Position: Professor

Printed Name: YUJI NAKAMURA
Organization: Tokyo Keizai University
Address: Telephone
No:
1-7-34 Minami-cho
Kokubunji-shi Date: March 4, 1998
Tokyo 185 Japan

III. Document Availability Information (from Non-ERIC
Source):

Complete if permission to reproduce is not granted to ERIC,
or if
you want ERIC
to cite availability of this document from another source.

Publisher/Distributor:

Address:

Price per copy: Quantity price:

IV. Referral of ERIC to Copyright/Reproduction Rights
Holder:

If the right to grant reproduction release is held by someone
other than the
addressee, please complete the following:

Name:

Address:

V. Attach this form to the document being submitted and
send or
fax to:

Acquisitions Coordinator
ERIC/CLL
1118 22nd Street, NW
Washington, DC 20037
FAX: 202-659-5641
TEL: 202-429-9292