

アンケート調査によるスピーキングテストの評価基準：モノログ、ダイアログ、マルチログ

Rating Criteria for the Three Speaking Test Format: Monologue, Dialogue and Multilogue

中村 優治 NAKAMURA, Yuji

● 慶應義塾大学
Keio University

Keywords スピーキング能力測定, 双方向対話, モノログ, ダイアログ, マルチログ, 評価基準
testing oral skills, interactivity, monologue, dialogue, multilogue, rating criteria,

ABSTRACT

本稿はスピーキング能力を言語能力テストの構成概念の考え方を基に、3つの側面（モノログ、ダイアログ、マルチログ）から捉え、アンケート方式により中学、高校、大学の先生が評価項目としてどのようなものを取り上げ、どの程度の相対的重み付けを行うのかを調査研究したものである。結果として、年齢が低い学年を教えている教員ほど、態度を重視していることが明らかになった。文法、語彙、流暢さは大学でより重視されており、また、説得性に関してはどのレベルでも等しく重要と考えられていることがわかった。この分析結果が評価の実際の中で評価者の評価行動にどのように反映されているのか、その実態の解明が継続課題となった。

A fundamental aspect of testing oral skills is the issue of interactivity. Effective communication is founded on one speaker having an effect on another, and on the reactions and responses between interlocutors. Each test format requires a great range of interactive skills such as delivering a monologue, participating in a dialogue, or multilogue (three way discussion). For raters, specific rating criteria should be taken into consideration because it is important to assess each skill correctly. The purpose of the paper, based on questionnaire results, is to assess the similarities and differences of rating criteria of student speaking skills among junior high school, high school and university level English teachers. The results indicate that attitude is seen as an important criterion at lower educational levels. Grammar, vocabulary and fluency are more weighted at university. However, persuasion, which seems difficult to measure, is equally important regardless of the level. Future study should examine how

these results are reflected in the raters' real rating.

1. Introduction

A fundamental aspect of testing oral skills is the issue of interactivity. Effective communication is founded on one speaker having an effect on another, and on the reactions and responses between interlocutors. (cf. Hughes, 2006). Each test format demands a great range of interactive skills such as delivering a monologue, participating in a dialogue, or a multilogue (three way discussion).

Nakamura (2005) suggested a framework by combining Nakamura's three-dimensional construct and Brown's (2004) five-category framework as follows:

<u>Nakamura</u>		<u>Brown</u>
Monologue	Less creative monologue (pre-monologue)	Imitative Intensive
	Creative Monologue	Extensive
Dialogue	Transactional dialogue Interpersonal dialogue	Responsive Interactive
Multilogue		Interactive

In this suggested framework, the term *Monologue* encompasses a range of one-way speaking proficiencies from a simple phonetic level check (Pre-Monologue) to advanced oral presentation skills (Creative Monologue). Dialogue, of course, refers to two-way, reciprocal speaking ability, but in this framework we distinguish two sub-types: *Transactional*, the ability to exchange information in situations where the context is well defined and with fixed conventions of speech e.g. classroom

situation; and *Interpersonal*, the ability to perform in consideration of deeper or less predictable social relationships. *Multilogue* refers to a speaker's discussion ability among multiple participants, not only as a participant but also in such more formalized roles as chairperson.

Students should be urged to develop oral communication-oriented learning habits involving the three-dimensional speaking construct discussed in this report since fluent oral communication in the target language is the ultimate goal for most, if not all language learners.

As Luoma (2004) indicates, the upcoming challenge with pair and group tasks is finding the types of tasks that are most appropriate for testing them. While pair and group tasks have been widely used in educational settings, more experiments to improve them will need to be conducted. The objective is to observe individual learning performance and analyze the strengths and weaknesses of different task features in particular assessment contexts. To ensure fairness of evaluation, the number of examinees a rater can observe in one interaction is to be determined. Further development of rating procedures for pair-group tasks is also needed. Further research should be carried out to solve these problems.

2. Background and Rationale

Luoma (2004) makes suggestions in her statement on "future directions in assessing speaking." First, in addition to communication-oriented tasks, communication-oriented criteria should be created, accordingly. Secondly, and more importantly, she introduces an idea of pair and group tasks for two reasons. Since speaking assessments are time-

consuming to administer and rate, efficient use of live testing and scoring time is to have two or more examinees interact with each other. In effect, this reveals aspects of their individual speaking ability and peer interaction in a context more realistic than interaction between interlocutor and an examinee. While in peer interaction the examinees relate to each other on equal grounds, in an interaction between an examinee and an interlocutor, the latter most likely holds power. That being said, as Luoma indicates, even in the case of peer interaction, the test discourse rests in the test developer's hands.

Brown (2004) identifies five categories of speaking performance assessment tasks: 1) imitative, 2) intensive, 3) responsive, 4) interactive, and 5) extensive. Accordingly, imitative assessment tasks should be used to check the test takers' phonetic level of production. No inferences are made about the students' ability to understand or convey meaning or to participate in an interactive conversation. Intensive assessment tasks concern the students' competence within a narrow band of grammatical, phrasal, lexical or phonological relationships. The test takers should be aware of semantic properties, with limited interaction with the test administrator. Possible tasks should ideally include direct response tasks, reading aloud, and sentence and dialogue completion. Extensive monologue assessment tasks include speeches, oral presentation and story-telling, where oral interaction from listeners is highly limited. It seems that the extensive assessment task should fit better as a responsive task in terms of the degree of interactional attribute. Responsive assessment tasks include interaction with limited to brief conversations, standard greetings and simple requests or comments. Interactive assessment tasks, which are basically similar to the responsive kind, are longer, more complex, and they include multiple exchanges and multiple participants. According to Brown (2004), interaction in such terms can take two forms. One is

Transactional, which consists in exchanging specific information, and the second is Interpersonal, by which social relationships are maintained. As Brown (2004) points out, oral production can become pragmatically complex in interactional exchanges.

The concept of multiple participants and the degree of interaction among participants is the primary concern both in Luoma (2004) and Brown (2004), and was also explored by Nakamura (1999, 2001, 2004). Dealing with speaking skill testing in a wider perspective, the latter took a three-dimensional approach (monologue, dialogue and multilogue) specifically focusing on the number of people involved in the process.

Van Moere further makes the observation that in group oral tests the rater observes and assesses each individual in the group on their own merits even while they are in interaction with their fellow candidates. The advantages of this test format over the more traditional interview tests are as follows (Van Moere, 2006):

- 1) It is a resource-saving way of assigning speaking scores to larger numbers of candidates as raters can test up to five or six examinees in one session.
- 2) It is less of a burden for examiners, as they are free to concentrate on candidates' performance without having to lead the conversation or refer to an interlocutor frame.
- 3) It may also encourage positive washback to the classroom in cultures where more communicative, speaking-focused teaching and learning goals are needed.
- 4) It links assessment to common classroom practice: when so much learning takes place as a result of small group collaboration and cooperation, it is arguably fairer for students to recreate similar conditions when assessing them.
- 5) Test takers have given positive reactions to this test format. They have reported to find

it less intimidating to hold a discussion with several peers rather than with an examiner in a one-on-one interview, to feel having more control over the direction of the conversation, and to be allowed to produce more natural language than they otherwise would.

3. Purpose of the Study

The purpose of the paper, based on questionnaire results, is to assess the similarities and differences of rating criteria of student speaking skills among junior high school, high school and university level English teachers. It is hoped that a dynamic assessment of the oral proficiency from three aspects (Monologue, Dialogue and Multilogue) can be explored.

4. Method

Subjects (N=104)

junior high school English teachers (n=49)

high school English teachers (n=45)

university English teachers (n=10)

Instruments

A questionnaire format (see Appendix).

Procedures

Subjects were asked to fill out a questionnaire.

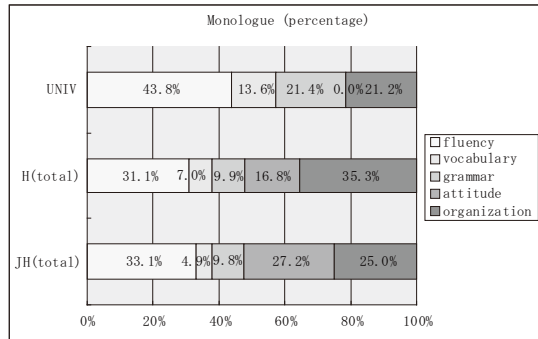
Analyses

The data was analysed using the Excel analysis program and the SPSS statistical program.

5. Results

Overall Tendency						
Monologue	Sampling	fluency	vocabulary	grammar	attitude	organization
	n					
JH	49	33.1%	4.9%	9.8%	27.2%	25.0%
H	45	31.1%	7.0%	9.9%	16.8%	35.3%
UNIV	10	43.8%	13.6%	21.4%	0.0%	21.2%
Dialogue	Sampling	fluency	vocabulary	grammar	attitude	comprehension
	n					
JH	46	32.6%	3.4%	7.9%	28.0%	28.1%
H	42	32.5%	4.4%	7.3%	11.0%	44.8%
UNIV	9	44.4%	16.4%	22.5%	1.4%	15.3%
Multilogue	Sampling	fluency	vocabulary	grammar	attitude	persuasion
	n					
JH	39	20.1%	3.4%	6.3%	31.8%	38.5%
H	38	19.4%	5.0%	7.0%	21.7%	46.9%
UNIV	9	28.7%	16.5%	17.9%	3.5%	33.4%

Table 1
Monologue (percentage)



[Monologue]

a. Fluency

fluency, pronunciation, intonation, speed, promptness, pause

b. Vocabulary

vocabulary, accuracy

c. Grammar

grammar, accuracy

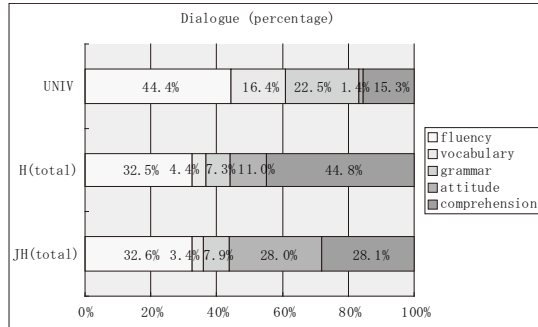
d. Attitude

attitude, loud/clear voice, eye contact, gesture/body-language, expression, delivery, mood, effort

e. Organization

organization, coherence, consistency, quantity, length (per 1 sentence), persuasion, memorization, clarity, originality, content, accuracy, emotion, clear concept of the text, topic, time constraints, bonus

Table 2
Dialogue (percentage)



[Dialogue]

a. Fluency

fluency, pronunciation, intonation, speed, natural/quick response, pause, timing

b. Vocabulary

vocabulary, accuracy

c. Grammar

grammar, accuracy

d. Attitude

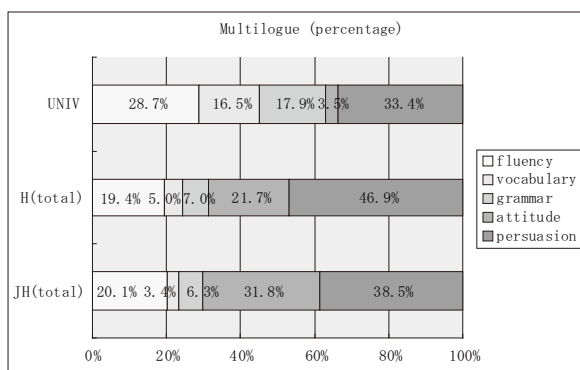
attitude, loud/clear voice, eye contact, active

participation, gesture/body-language, cooperation, participation, feeling, expression, delivery,

e. Comprehension

comprehension, organization, ask and answer properly, appropriateness, persuasion, clarity, engaged listening, quantity, content, accuracy, communication ability, function, clear the task, understanding the question, discourse competence, dialectic, having cultural background and knowledge

Table 3
Multilogue (percentage)



[Multilogue]

a. Fluency

fluency, pronunciation, intonation, speed, natural/quick response, pause, tone of voice, promptness

b. Vocabulary

vocabulary, accuracy

c. Grammar

grammar, accuracy

d. Attitude

attitude, leadership, participation, cooperation, loud/clear voice, eye contact, gesture/body-language, aggressive, positive participation, not shy, teamwork, feeling, expression, delivery, number of turns, plus more answers, playing a required role, making a friendly atmosphere, respecting partners, "May I speak...?"

e. Persuasion

persuasion, consistency, organization, ask and answer properly, appropriateness, logical argument, quantity, content, accuracy, communication ability, function, expressing ideas clearly, task completion, listen to the partner, hearing, understanding partner's opinion, say the opinion oneself, having cultural background and knowledge

6. Overall Discussion

1. The lower the educational levels are, the more

important attitude becomes. Presumably it is taken for granted that attitude is already a proviso for college students.

2. In the Monologue and Dialogue skills, Fluency including pronunciation is regarded as important among all of the three levels. This is especially true of university teachers. In Multilogue, Fluency is rated relatively lower when compared with Monologue and Dialogue.
3. At the university level, Vocabulary and Grammar are given twice as much weight compared to junior and senior high schools. Between junior high school and senior high school teachers, the latter insist more on vocabulary and grammar.
4. The last (fifth) category in each skill, i.e. Organization in Monologue, Comprehension in Dialogue, Persuasion in Multilogue, are the most highly thought of by high school teachers, followed by junior high school and university teachers. Although it is very difficult to establish a clear rating criteria for these three categories (organization, comprehension and persuasion), the teachers among all of three education levels commonly ask for the same three things: 1) easy comprehension of the assigned task for students, 2) competency of speech organization, 3) easy comprehension of the counterparts and their appropriate responses to them.

7. Conclusions and Implications

The results of the questionnaire in this study demonstrate a way of looking at the oral skills from three tasks (Monologue, Dialogue and Multilogue).

According to Van Moere (2006), most of the performance-based oral testing research literature has focused on the traditional face-to-face interview or simulated (semi-direct) tape mediated interviews. However, the need for language testers to control the test situation, elicit ratable language samples in allocated segments of time, or create scenarios to force particular language functions may have caused them to overlook opportunities where examinees can be observed producing rich, natural language in conversation with their peers.

As Hughes (2002) underlines, a fundamental aspect of testing oral skills is the issue of interactivity. Good communication is founded on one speaker actually having an effect on another, and on the reactions and responses which take place between interlocutors. The Certificates in English Language Skills (CELS) test format and focus for assessment is an example of a test framework which takes into account far more of the dynamic and interpersonal facets of speech communication (Hughes, 2002). This format demands a greater range of interactive skills to be used (individual speech, dialogue and three-way discussion) and a variety of spoken genres and registers to be produced.

Although it requires time and effort to establish clear rating criteria for different categories in individual tasks, there is also an increased reliability risk of test scores. Therefore, we need to consider the potential advantages of group assessments: the benefit of being able to observe examinees in natural interaction with one another, and allowing them to direct and control their own discourse (cf. Van Moere, 2006).

Acknowledgement

I wish to thank Mr. Takanori Yano, a Keio University student, for his support and help in the analysis of this data.

References

- Bachman, L.F. and Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bachman, L. F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition* 10, 149-64.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing* 20,1,1-25.
- Brown, H.D. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education.
- Canale, M. & M. Swain. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1) 1-47.
- Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language*. Cambridge: CUP.
- Fulcher, G. (1996). Testing tasks; issues in task design and the group oral. *Language Testing* 13, 23-51.
- Fulcher, G. (2003). *Testing Second Language Speaking*. London: Pearson Education.
- Hawkey, R. (2004). A Modular Approach to Testing English Language Skills The development of the Certificate in English Language Skills (CELS) examinations: *Studies in Language Testing* 16. Cambridge, UK: Cambridge University Press.
- He, A. W. & Young, R. (1998). Language proficiency interviews: A discourse approach. In Young, R. and He, A.W. (eds) *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency*. Amsterdam: John Benjamins.
- Hilsdon, J. (1991). The Group Oral Exam. Advantages and Limitations. In J.C. Alderson and B. North (eds). *Language Testing in the 1990s*. London: Macmillan. (pp.189-197).
- Hughes, R. (2002). *Teaching and Researching Speaking*. Pearson Education; London.
- Hymes, D. (1972). On communicative competence. In J.B. Pride and J. Holmes (eds.), *Sociolinguistics*. Harmondsworth: Penguin, pp. 269-293.
- Leung, C. (2004). Developing Formative Teacher Assessment: Knowledge, Practice, and Change. *Language Assessment Quarterly*, 1,1, 19-41.
- Analysis*. Chicago, IL: MESA Press.

- Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- Lynch, B. K. (2003). *Language Assessment and Programme Evaluation*. Edinburgh: Edinburgh University Press.
- McNamara, T. (1996). *Measuring Second Language Performance*. London: Longman.
- Nakamura, Y. (1993). Measurement of Japanese College Students' English Speaking Ability in a Classroom Setting. Unpublished Ph. D Dissertation. International Christian University.
- Nakamura, Y. (1999). Measuring Speaking Skills through Multidimensional Performance Tests. *Educational Studies*, 41, 99-113. International Christian University.
- Nakamura, Y. (2001). Rasch Based Analysis of Oral Proficiency Test Data. *Educational Studies*, 43, 191-197. International Christian University.
- Nakamura, Y. (2003). Oral Proficiency Assessment: Dialogue Test and Multilogue Test. *2003 JALT Pan-SIG Conference Proceedings*. 51-61. The Japan Association for Language Teaching: Tokyo. Online Available at <http://www.jalt.org/pansig/2003/HTML/Nakamura.htm>.
- Nakamura, Y. (2004). A Many-Facet Rasch Analysis of Three-Dimensional Oral Proficiency Test Data. *Communication Studies* 20, 119-142. Tokyo Keizai University.
- Nakamura, Y. (2006). The Construct of Speaking for Communicative Testing Journal of Foreign Language Education Vol.2 pp.61-74 Keio Research Center for Foreign Language Education.
- O'Loughlin, K. (2001). The equivalence of direct and semi-direct speaking tests: *Studies in Language Testing* 13. Cambridge, UK: Cambridge University Press.
- Reves, T. (1991). From Testing Research to Educational Policy: A Comprehensive Test of Oral Proficiency. In J.C. Alderson and B. North (eds). *Language Testing in the 1990s*. London: Macmillan.(pp.178-188).
- Shohamy, E. (1983). The stability of the oral proficiency trait on the oral interview speaking tests. *Language Learning* 33,527-39.
- Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing* 11, 99-124.
- Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing*, 23, 4, 411-440.

Appendix

Questionnaire used for this paper

1. Test tasks of a speaking test

1). Monologue

(one-way speaking, no listening ability is required, a speech making test or a picture description test: ex. speech)

2). Dialogue

(two-way speaking, intermediate listening ability is required, a live interview test or a tape mediated speaking test: ex. interview)

3). Multilogue

(more than three people are involved, advanced listening ability is required, discussion or debating: ex. debate)

2. Rating Items of each task

1). Monologue

(ex. Speech: pronunciation, fluency)

2). Dialogue

(ex. Interview: prompt reply, appropriateness)

3). Multilogue

(ex. Debate: participation, persuasion, logicity)

3. Weighting of rating items

- 1) Are all rating items equally important?
- 2) How is the degree of importance of each rating item decided?
- 3) Examples:

(1). Monologue

(ex. Speech: Fluency, Pronunciation, Grammar, Vocabulary)

5 2 2 1

(2). Dialogue

(ex. Interview: Fluency, Pronunciation, Grammar, Vocabulary)

4 2 2 2

(3). Multilogue

(ex. Debate: Persuasiveness, Pronunciation, Grammar, Vocabulary)

5 1 2 2

4. Please decide rating items for each task (monologue, dialogue, multilogue) and give a weighting to each item as examples mentioned above. The total ratio is 10.