A Comparative Analysis of Semi-direct Speaking Testing and Direct Speaking Testing for Japanese EFL Learners

Michi YONEZAKI

Although speaking tests have become more important in Japan over the past decades with the recent emphasis on communicative language teaching, they remain infrequent at most schools. This is because assessing speaking skills is quite difficult in terms of its practicality. However, since MEXT (2014b) announced that students’ English proficiency and learning status in terms of the four language skills will be assessed and analyzed even in entrance examination in the near future, speaking test development can be on urgent business. There are two main types of test formats to assess speaking skills so far: a semi-direct, or person-to-machine, test, and a direct, or face-to-face, test. This study investigated whether the speaking ability measures would make a difference depending on two test formats. The results showed there was no statistically significant difference between two test formats. However, according to questionnaire results, test takers preferred direct testing to semi-direct testing.

Key words: Speaking Test, Semi-direct Testing, Direct Testing

1. Background

1.1. The Situation of Assessment of Speaking Tests

Speaking tests have become more important over the past decades with the recent emphasis on communicative language teaching in Japan. According to survey results of English education for 2014 by Ministry of Education, Culture, Sports, Science & Technology (MEXT, 2014a), 50.4% schools (general course) in the survey said that they conducted a performance test, such as speaking tests and writing tests, in Communication English I, 44.3% in Communication English II, 55.2% in English Expression I, and 52.6% in English Expression II. The ratios became much better compared to the previous survey conducted by MEXT (2011)1. However, this performance test included writing tests, which are considered, in general, is easier to carry out. In addition, if only one speaking test was conducted in a semester or even in a year, the school can say that they conducted the performance test in this survey. Therefore, we cannot see whether the ratio of this result will be satisfactory or not.

Although more teachers have attempted to conduct speaking tests in English class, the frequency of the speaking test seems to be still low at school in Japan. A measure of how low the frequency of testing is can be seen in Hirai and Koizumi (2009) reporting of Honda’s (2007) survey that the teachers at lower secondary schools in Japan evaluated
the speaking performance of their students only 1.2 times per semester. They also predicted, in the same article, a much lower frequency of speaking assessment at upper secondary schools.

In general, teachers face many obstacles to assess a speaking test. One of the obstacles is that a speaking test requires teachers to take significant amount of time in order to conduct and to score it. Tremendous burden on teachers results in a low practicality of carrying out speaking test. Hirai and Koizumi (2009) argue that the students do not take speaking activities seriously unless teachers do a speaking assessment in English class even though they give their students opportunities for speaking activities.

Another obstacle which teachers, in particular, at upper secondary school face is little speaking assessment in university entrance examinations (Hirai & Koizumi, 2009; Shirai, 2012). In order to alleviate this problem and also to cultivate students' communicative competence in English, MEXT (2014b) announced that upper secondary schools and universities will assess and analyze students' English proficiency and learning status in terms of the four language skills and utilize the results. It also announced even in entrance examinations, communication skills in terms of the four language skills need to be evaluated. This means that learners' speaking skill, which seems to have been ignored, will be more emphasized and measured with other skills. When the current second year students at lower secondary school become third-year students at upper secondary school, they are supposed to take an entrance examination including external standardized tests, which integrates four language skills. This is a commendable move, but speaking test development is still a work in progress in Japan. Therefore, it is on urgent business and should be a target for future research and development. For the further development of speaking tests, however, there is a need to evaluate speaking test formats and test takers' ability measures.

1.2 Formats for Speaking Assessments

There are three types of test formats to assess speaking skills: indirect testing, direct testing, and semi-direct testing (Hughes, 2003; O'Loughlin, 2001; Qian, 2009). Of these testing formats, indirect testing is viewed as the least valid measure of speaking ability because test takers are not required to speak (O'Loughlin, 2001). Semi-direct testing and direct testing are general viewed as legitimate speaking testing formats. Therefore, this study focuses on direct testing and semi-direct testing.

In semi-direct testing, test takers respond to audio stimuli or directions written on paper, and their own responses are recorded and later assessed. Semi-direct testing can be carried out to a large number of test takers at the same time and within a very short period as long as there are facilities available, such as language laboratories or IC recorders, to record test takers' performance. Therefore, the advantage of semi-direct testing is that it is cost-effective and efficient (Qian, 2009). With the development of computer technology, the semi-direct test has become more popular. Many international oriented English language proficiency tests, such as the Test of English as a Foreign Language Internet-based test (TOEFL iBT test) and TOEIC Speaking Test, also have a semi-direct speaking test format with the computer-based version.

On the other hand, in a direct speaking test, test takers are required to perform oral tasks which can show their speaking skill and to involve face-to-face communication with one or more interviewers. Examples of this type of test are EIKEN speaking tests, and the International English Language Testing System (IELTS), which has been developed by the University of Cambridge Local Examination Syndicate. Direct testing reflects
more authentically face-to-face interaction in daily life. Therefore, direct testing is believed to offer a much better test validity (Qian, 2009). However, interviewers control the interview and "the language elicited in an interview is unlikely to reflect the discourse of real-life conversation (O'Loughlin, 2001, p.6)."

1.3. Previous Research Findings
In comparing these two speaking testing formats, a number of researches have been conducted in various geographical regions by analyzing and comparing the features and structures, and test takers’ reactions to and perceptions of semi-direct tests and direct tests (e.g., Brown, 1993; Deguchi, 2013; Hughes, 2003; Luoma, 2004; Nakamura, 2015; O’Loughlin, 2001; Qian, 2009).

Table 1 shows a list of the results drawn from the literature of test takers’ reactions to test formats. Although the studies on test-takers’ preference have produced various results, most studies reported that test takers, in general, prefer direct tests for speaking tests. However, this does not mean semi-direct tests will always be unpopular. Brown (1993) reported that 57% of the test takers liked a semi-direct test, which was a tape mediated test of spoken Japanese for the tourism and hospitality industry in Australia. In her study, only 25% test takers said that they disliked a semi-direct test and 18% were neutral. One of the reasons for the positive reaction to a semi-direct test is that this test was a specific purpose test context and it reflected the type of language required of employees in the tourism and hospitality industry. Therefore, if the test takers had been given an opportunity to take the same test with a direct test format, they might have showed the positive reaction to a direct test as well.

O’Loughlin (2001) conducted two trials with the direct and semi-direct test formats of oral interaction sub-test in 1992 and 1994. Although questions in both test formats were as close as possible, the test takers preferred the direct testing. This result was the same

<table>
<thead>
<tr>
<th>Study (research year)</th>
<th>Test Type (or name)</th>
<th>Candidates (N)</th>
<th>In Favor of Direct Test</th>
<th>In Favor of Semi Direct</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown (1993)</td>
<td>Occupational foreign language test (OFLT)</td>
<td>University students and Adults in Japanese courses (53)</td>
<td>–</td>
<td>57%</td>
<td>–</td>
</tr>
<tr>
<td>Nakamura (2015)</td>
<td>Oral-English proficiency test: eiken</td>
<td>Japanese EFL learners (109)</td>
<td>53.4%</td>
<td>31.1%</td>
<td>15.5%</td>
</tr>
<tr>
<td>O’Loughlin (1992)</td>
<td>Oral-interaction sub-test: access</td>
<td>NESB volunteers from language centers and tertiary institutions (94)</td>
<td>90%</td>
<td>10%</td>
<td>–</td>
</tr>
<tr>
<td>O’Loughlin (1994)</td>
<td>Oral-interaction sub-test: access</td>
<td>NESB volunteers from language centers and tertiary institutions (77)</td>
<td>94%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>Qian (2009)</td>
<td>Oral-English proficiency test</td>
<td>the Hongkong Politechnic University students (243)</td>
<td>32.8%</td>
<td>9.7%</td>
<td>57.6%</td>
</tr>
</tbody>
</table>

NESB = non-English speaking background
with Nakamura’s (2015) study, which compared a newly developed computer-based speaking test to a face-to-face speaking test in EIKEN. O’Loughlin (2001, p.143) reported that semi-direct tests were comprised entirely of “a set of one-way remote exchanges in which the candidate prepares and responds within rigid time limits to an audio recorded voice.” This might be the reason for some negative reactions to semi-direct testing. He determined that direct tests, on the other hand, involved a two way personal interaction between test takers and interviewers. This is a typical reason given that direct testing was more preferable.

On the other hand, Qian (2009) reported that test takers had no particular preference. Unlike other researches, Qian prepared two different oral English proficiency tests. This is because he thought that the two testing formats did not necessarily tap the same type of skills, and his main concern was to investigate how test takers would view the two formats of direct testing and semi-direct testing. The IELTS Speaking sub-test was employed for a direct test while Graduating Students’ Language Proficiency Assessment-English (GSLPA), a performance-based English proficiency test developed at the Hong Kong Polytechnic University, was used for a semi-direct test. The employment of different oral tests in each test may have caused the results.

Although there is some research on test takers’ reactions toward testing formats, the relationship between test scores and test formats has not been examined except Nakamura’s (2015) research. Little is known about whether there is a difference on speaking test scores between the two test formats. Therefore, there is a need for research to understand the test takers’ measure of ability. In addition, although assessing speaking skills with computers or other formats of multimedia technology has been popular in recent years, we need to listen to the voices from the test takers’ perspectives and then investigate the test formats for future development.

In this study, direct testing is defined as live, face to face testing where test takers and one or more interviewers are on the spot, while semi-direct testing is defined as tape or digital-recorded testing in which test takers respond to audio-recorded questions or to the directions on printed paper.

2. Experiment

2.1 Purpose

The aim of this study is to examine in terms of the comparability study of a direct speaking test and a semi-direct speaking test. This study also examines test takers’ reactions to and perceptions of direct testing and semi-direct testing.

The specific research questions that the study aims to investigate are as follows:
1. Is there a significant difference in the test scores between a direct test and a semi-direct test?
2. How do test takers react to and perceive each of the test formats?

2.2 Participants

Forty-three students learning English as a foreign language were recruited in this test as test takers. All of them are freshmen at a Japanese university and the researcher teaches them English. The average English participants’ proficiency levels were Level 3 for ELPA (Association for English Language Proficiency Assessment) placement test, which they took in April, 2015. Level 3 for ELPA placement test equals to the pre-second
A Comparative Analysis of Semi-direct Speaking Testing and Direct Speaking Testing for Japanese EFL Learners

grade level for the Society for Testing English Proficiency (STEP), which means that the participants' proficiency levels are considered as low to intermediate.

2.3. Procedures

Due to the limitation of the available facilities, this study used IC recorders to record the participants' performance in semi-direct testing. Most of the participants didn't know how to use an IC recorder. Therefore, before the tests were conducted, the participants recorded their performance (retelling the story) on an IC recorder so that they might not be confused about how to use an IC recorder in testing.

In addition, many participants were not used to taking speaking tests. Therefore, the first speaking test format (either a direct or a semi-direct test) is likely to make them more nervous than the second speaking test format. In order to make the two test formats equivalent, the participants were divided into two groups (Group 1 and Group 2) randomly.

Group 1 took a direct test first and then took a semi-direct test. On the other hand, Group 2 took a semi-direct test first and then took a direct test. All the participants took both a direct test and a semi-direct test in different order in an English class.

In both tests, the participants were not allowed to bring anything. In a direct test, the interviewer (the author) conducted a face-to-face interview with each test taker. The interviewer gave oral directions in English to the participants. If they did not understand the interview's direction, they were allowed to ask the questions again. In a semi-direct test, the participants recorded their performance on IC recorders following the directions written on a paper. The participants were not allowed to record their speaking on the IC recorder again.

2.4. Test Materials and Test Formats

As a speaking assessment, there are two parts of the speaking test in both tests. The first part is the Story Retelling Speaking Test (SRST), developed by Hirai and Koizumi (2009). In the SRST, the participants were required to read a story and retell it. In Hirai and Koizumi's research, test takers were required to read a story which they read for the first time. However, this study asked the participants to read a story which they had already learned in an English lesson and then retell it. This is because some of the participants' proficiency levels were considered as low. In addition, in daily English classes, story retelling activities were not conducted. Therefore, it might be difficult for some of the participants to read a new story and retell it within a given time. It is likely that some of the participants could not manage to retell a new story. Therefore, this study used the story they had already learned although doing this may impact on the results.

Two types of stories (Story A and Story B) from the textbook, New Headway Beginner (Oxford), which the participants learned in English lessons, were employed for the tests. Readability for Flesch reading ease test is 98.3 for Story A and 93.5 for Story B. The level of both stories is elementary and both stories are the same level.

In the second part of the speaking test, the participants were required to state their opinions after retelling the story. The questions were used in both tests. The questions are as follows:

1) Did you like this story? If you like this story, tell me why. If you don't, why not?
2) Which character do you like the best in this story? Tell me the reason.
As mentioned above, both groups (Group 1 and Group 2) took a direct test and a semi-direct test. In order to make the two test formats equivalent, Group 1 took a direct test first with Story A and then took a semi-direct test with Story B. On the other hand, Group 2 took a semi-direct test first with Story B and then took a direct test with Story A. Table 2 shows the outline of the experiment.

Table 2. The Outline of the Experiment

<table>
<thead>
<tr>
<th>Participants</th>
<th>The first speaking test</th>
<th>The second speaking test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Direct test</td>
<td>Semi-direct test</td>
</tr>
<tr>
<td></td>
<td>(SRST + stating opinions)</td>
<td>(SRST + stating opinions)</td>
</tr>
<tr>
<td>Group 2</td>
<td>Semi-direct test</td>
<td>Direct test</td>
</tr>
<tr>
<td></td>
<td>(SRST + stating opinions)</td>
<td>(SRST + stating opinions)</td>
</tr>
</tbody>
</table>

The participants were informed of which stories of the textbook they were supposed to read and retell beforehand (Lesson 1 and Lesson 2 in *New Headway Beginner*, Oxford). Therefore, some participants prepared for the speaking test at home. However, they were not informed of the questions which they would be asked after retelling. The participants were not allowed to bring anything such as pencils, pens, or handouts during the tests. They were not allowed to write and they were told to retell orally.

2.5. Questionnaire

To examine about the participants' affective reactions to the direct test and the semi-direct test, two-part of the questionnaire was conducted. The participants answered questions using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The first-part items are as follows;

1) I got nervous during the test (nervousness).
2) I thought the test was difficult (the test difficulty).
3) I thought that I had the favorableness for the test (the test favorableness).
4) I performed well (the test performance).
5) I thought the test was useful (the test usefulness).

In the second-part, the study inquired about the participants’ preference of the two test formats. The participants were asked to choose a direct test, a semi-direct test or no preference and then to write their opinions freely.

3. Results

3.1. Method of Analysis

In order to examine direct speaking testing and semi-direct speaking testing, in terms of the comparability, the results of each speaking test were analyzed. In total, forty-three students speaking performances of the two test formats were analyzed.

The performances of the participants were evaluated against the four criteria: volume, content, fluency and grammatical accuracy as indicated in Appendix. The author evaluated the speaking performance of the participants. Five levels of performance (1 to 5) were prepared for each of the four criteria. The total score is 40 points (two parts of
the speaking tests of five points for four criteria).

3.2. Comparison between a Direct Speaking Test and a Semi-Direct Speaking Test

Table 3 provides the descriptive statistics of direct testing and semi-direct testing scores. It provides the mean scores ($M$) and standard deviations ($SD$) for each speaking test part (the first part and second part) and the total of the speaking test.

| Test format | $n$ | First part | | | Second part | | | Total | | |
|-------------|-----|------------|-----|------|------------|-----|------|-----|------|
|             |     | $M$   | $SD$ | $M$   | $SD$ | $M$   | $SD$ |       |       |
| Semi-direct | 43  | 11.56 | 2.13 | 9.79  | 1.51 | 21.35 | 3.24 |       |       |
| Direct      | 43  | 11.90 | 2.06 | 9.40  | 1.76 | 21.30 | 3.11 |       |       |

In order to examine whether there is a significant difference between two-test formats, paired $t$-test was conducted. The difference of the total scores of the speaking test between two tests was not statistically significant ($t (84) = 0.07, p = .95, d = 0.02$). Nor was the statistically significant difference of each speaking test part found (the first part of speaking test: $t (84) = -0.77, p = .44, d = -0.17$; the second part of the speaking test: $t (84) = 1.12, p = .27, d = 0.24$).

3.3. Further Analysis

Although there was no statistically significant difference between a semi-direct test and a direct test, the study tried to figure out whether there would be a difference on each criterion for each part of the speaking test.

Table 4 presents the mean values ($M$) and standard deviation ($SD$) for a semi-direct test and a direct test in each criterion.

<table>
<thead>
<tr>
<th>Speaking test criteria</th>
<th>Semi-direct ($n = 43$)</th>
<th>Direct ($n = 43$)</th>
<th>$t (84)$</th>
<th>$p$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>volume</td>
<td>$M$ 5.00 $SD$ 0.98</td>
<td>$M$ 4.74 $SD$ 0.98</td>
<td>$t$ 1.21</td>
<td>.23</td>
<td>0.27</td>
</tr>
<tr>
<td>content</td>
<td>$M$ 5.51 $SD$ 1.03</td>
<td>$M$ 5.49 $SD$ 0.91</td>
<td>$t$ 0.11</td>
<td>.91</td>
<td>0.02</td>
</tr>
<tr>
<td>fluency</td>
<td>$M$ 5.33 $SD$ 1.11</td>
<td>$M$ 5.54 $SD$ 1.03</td>
<td>$t$ -0.91</td>
<td>.37</td>
<td>-0.20</td>
</tr>
<tr>
<td>accuracy</td>
<td>$M$ 5.51 $SD$ 0.91</td>
<td>$M$ 5.54 $SD$ 0.80</td>
<td>$t$ -0.13</td>
<td>.90</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

As Table 4 shows, the mean values of the semi-direct test on volumes and content were slight higher than those of the direct test, while the mean values of the direct test on fluency and accuracy were slight higher than those of the semi-direct test. However, there was no statistically significant difference for any criteria between the direct test and the semi-direct test (volume: $t (84) = 1.21, p = .23, d = 0.27$; content: $t (84) = 0.11, p = .91, d = 0.02$; fluency: $t (84) = -0.91, p = .37, d = -0.20$; accuracy: $t (84) = -0.13, p = .90, d = -0.02$).

3.4. Questionnaire Result

In order to examine the participants’ reactions to and perceptions of direct testing
and semi-direct testing, the questionnaire was issued soon after the participants took each speaking test. Table 5 provides the descriptive statistics concerning the results of the questionnaire.

Table 5. Questionnaire Result

<table>
<thead>
<tr>
<th></th>
<th>Semi-direct (n = 43)</th>
<th>Direct (n = 43)</th>
<th>t (84)</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>t</td>
</tr>
<tr>
<td>Nervousness</td>
<td>4.49</td>
<td>0.83</td>
<td>4.00</td>
<td>1.18</td>
<td>2.23</td>
</tr>
<tr>
<td>Test difficulty</td>
<td>3.51</td>
<td>1.10</td>
<td>3.35</td>
<td>0.95</td>
<td>0.74</td>
</tr>
<tr>
<td>Test favorableness</td>
<td>3.00</td>
<td>0.85</td>
<td>3.63</td>
<td>1.05</td>
<td>-3.06</td>
</tr>
<tr>
<td>Test performance</td>
<td>2.58</td>
<td>1.03</td>
<td>3.09</td>
<td>0.95</td>
<td>-2.40</td>
</tr>
<tr>
<td>Test usefulness</td>
<td>3.76</td>
<td>0.81</td>
<td>3.80</td>
<td>1.01</td>
<td>-1.06</td>
</tr>
</tbody>
</table>

*p < .05, ** p < .01

Unlike the test score results, statistically significant differences between the direct test and the semi-direct test were found on nervousness (t (84) = 2.23, p = .03, d = 0.49), the test favorableness (t (84) = -3.06, p = .00, d = -0.67) and the test performance (t (84) = -2.40, p = .02, d = -0.52). On the other hand, there are no statistically significant differences on the test difficulty (t (84) = 0.74, p = .46, d = 0.16) and test usefulness (t (84) = -1.06, p = .29, d = -0.23).

The study also asked the participants which test format they preferred. Table 6 shows the results. The number in favor of a direct test (n = 21, or 48.8%) greatly exceeded the number in favor of a semi-direct test (n = 9, or 20.9%), which indicates that the results of this study also tend to follow the results drawn from most of the previous research.

Table 6. Test Takers’ Favor of the Testing Format

<table>
<thead>
<tr>
<th>Testing format</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>In favor of a direct test</td>
<td>9</td>
<td>20.9</td>
</tr>
<tr>
<td>In favor of a semi-direct test</td>
<td>21</td>
<td>48.8</td>
</tr>
<tr>
<td>No opinion about either format</td>
<td>13</td>
<td>30.2</td>
</tr>
</tbody>
</table>

However, we should take note that 13 participants or 30.2% held no opinion about either testing format. In other words, about one-third of the participants did not show a particular preference in terms of the testing format.

In order to better understand why more participants preferred a direct test to a semi-direct test, the participants’ comments were also collected and analyzed. Some of the participants’ typical comments which were in favor of the direct test format are as follows:

- I prefer a face-to-face speaking test. The teacher nodded during the test even though I may have made mistakes, which made me more encouraged and less nervous.
- I enjoyed the speaking test more when I talked to someone who listened to me.
- I like talking to someone, not to an IC recorder.
- Talking to an IC recorder scared me because it showed no reaction.
A face-to-face speaking test, or speaking to someone, is more natural in our daily life. I like face-to-face speaking better because this test gave me a chance to talk with a live person.

These comments indicate that the participants wanted to have face-to-face interactions with humans in speaking testing, which reflects real life communication. On the other hand, there were also some participants who viewed semi-direct testing positively although the number of such comments was few.

I prefer speaking to an IC recorder because I feel more pressure to speaking English in front of someone.
I did not get nervous in semi-direct testing because I carried the speaking test out by myself. I didn’t have to care about the examiner.
I spoke at my own space and nobody listened to me in semi-direct testing.

Some of neutral opinions are as follows:
I got nervous in both tests.
I enjoyed both tests.
I felt worried in both tests because I did not know my answer was correct or not.

Putting the questionnaire results together, it can be said that the participants were likely to have more positive feelings for a direct test than a semi-direct test although there was not statistically significant difference found between the two speaking test scores.

4. Discussion

Findings from the results indicate three things. First, although there have been a debate over the appropriateness of two different testing formats (Qian, 2009), the results of this study found out that there was no statistically significant difference between two testing formats in terms of test scores. This result conforms to Nakamura's (2015) study. The result of this study may make test developers or teachers, who have tried to develop test formats equivalent in terms of test takers' speaking ability measurement, somewhat comforted. Although further empirical evidence is needed in order to be verified, this study led to one step to prove the equivalence of direct and semi-direct speaking testing.

Second, this may be somewhat contradictory to the first indication mentioned above. With the development of computer technology, semi-direct tests have become more popular. However, this study as well as most of the previous research found that test takers would view direct testing more favorably than semi-direct testing. If this is also true with the majority of test takers, test developers and teachers need to pay more attention to the test formats with the regard to the test takers' preference and keep in mind why they prefer direct testing. The main reason is, as some participants of this study pointed out, that direct testing allows test takers to have face-to-face communication with humans. This reflects the way most people would communicate in a real life. Talking to an IC recorder or a machine seems an artificial setting and does not reflect real
Finally, this study employed the SRST, which was developed by Hirai and Koizumi (2009) as a speaking test. In both speaking tests, the participants thought that employing speaking tests was useful in learning. In other words, the test takers showed a positive attitude toward the speaking test. Apart from what kind of speaking testing is appropriate, it is more important to give opportunities to speaking activities and assess students’ speaking ability in English class.

5. Conclusion, Limitation, and Implication for Further Research

This study investigated whether the speaking ability measures would make a difference depending on two testing formats, direct testing or indirect testing. This study also examines test takers’ affective reactions to and perceptions of direct testing and semi-direct testing. The results showed that there was no statistically significant difference between two test formats. However, according to the questionnaire results, test takers preferred direct testing to semi-direct testing and had more positive feelings toward direct testing.

As a pedagogical implication of this study, it can be said that the most ideal of a speaking test format is direct testing. However, in situations with a lack of human resources and time, semi-direct testing can be an alternative measurement of speaking ability. The more important thing is that teachers should give their students a speaking assessment once they let students involve speaking activities.

Several limitations of this study need to be pointed out. First, because of a human resource difficulty, this study’s examiner was the author, who was familiar to the participants. Therefore, further research is needed to investigate a direct test with different types of an examiner such as an examiner unfamiliar to test takers, a native English speaker and so on. As another lack of a human resource, the researcher scored the speaking tests. Although the scores were rated following a speaking scale, more raters were needed.

Second, this study employed the SRST, as a speaking test. Although most of the participants thought the SRST useful in learning, it is necessary to examine with another test material.

Finally, because of a technical difficulty at university where the researcher works, a computer-based testing form was not conducted. Instead of computers, IC recorders were used in a semi-direct test. In the future, if the facilities are available, we would like to conduct supplementary studies in order to confirm these findings.

Although several limitations exist, we hope that our findings will be a stepping stone to further and more detailed investigation of speaking test development.

Note

1. In the report on the result of curriculum organization and English education in public secondary upper school for 2010 (MEXT, 2011), 23.8 % schools (general course) said that they conducted a performance test in English I. The ratio was surveyed only about English I.
References


Appendix

Speaking Scales

<table>
<thead>
<tr>
<th>Volume</th>
<th>Content</th>
<th>Fluency</th>
<th>Grammatical accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>The volume is sufficient. Gives more than 70 words in total.</td>
<td>gives a clear description. Ideas logically presented and well structured.</td>
<td>expresses herself fluently and spontaneously, almost effortlessly.</td>
</tr>
<tr>
<td>4</td>
<td>The volume is enough. Gives 70~50 words in total.</td>
<td>gives a clear description. Ideas presented well.</td>
<td>produces language with a fairly good tempo. Although she searches for patterns and expressions, there are few long pauses.</td>
</tr>
<tr>
<td>3</td>
<td>gives 49~30 words in total and the volume is accepted.</td>
<td>gives a somewhat clear description but talks lack structure.</td>
<td>keeps going comprehensibly, even though pausing for grammatical and lexical planning and repair occurs sometimes.</td>
</tr>
<tr>
<td>2</td>
<td>gives 29~10 words in total and the volume is not enough.</td>
<td>Ideas not logically presented.</td>
<td>constructs phrases despite very noticeable hesitation and false starts. Very limited range of expression.</td>
</tr>
<tr>
<td>1</td>
<td>Gives only a few words (less than 10 words) or no statement.</td>
<td>No statement.</td>
<td>manages very short and isolated utterances with much pausing to search for expressions. Very limited range of expression.</td>
</tr>
</tbody>
</table>

Author
Michi YONEZAKI
Konan Women’s University
E-mail: michiyonezaki@gmail.com