

The use of oral proficiency tests in the
Japanese EFL context: Learners' perceptions

FUJII Akiko

WATANABE-KIM Izumi

IINO Atsushi

Abstract

The purpose of the study was to explore (1) Japanese EFL learners' performance on different types of oral proficiency assessments, and (2) Japanese EFL learners' perceptions of different types of oral proficiency assessments. Participants were 26 university students at a private university in Tokyo. They participated in (1) a face-to-face direct oral proficiency test interview based on the IELTS model, (2) a semi-direct recorded oral proficiency test in groups following the TOEFL iBT model, and (3) a self-assessment based on the CEFR model. Learners also completed questionnaires that elicited perceptions about the different assessment measures. Data consisted of learners' scores on the two types of tests (IELTS-type and TOEFL-type), learners' self-assessments, and questionnaire responses. Results indicated positive correlations between the IELTS-type and TOEFL-type tests but indicated negative correlations between the oral proficiency tests and the CEFR-type self-assessment. Learners' perceptions indicated that the tests raised their awareness of weaknesses in their language competence, such as vocabulary knowledge and listening comprehension. However, learners lacked clear views about what they could do to better prepare for the test. The study discusses the implications for the use of speaking tests within the EFL university context and suggests directions for further research.

Introduction

The primary function of foreign language testing has traditionally been for high-stakes selection or screening. In Japan, international high-stakes assessments such as the TOEFL iBT and the IELTS are often the largest challenges learners face when they wish to study abroad. Foreign language tests also play a decisive role in high school and university entrance screening, and the TOEIC test is used widely in the corporate sphere to screen job applications by university students. However, within the field of language testing, more and more attention has been directed to other more learner-oriented purposes of language testing such as diagnostic language testing, where tests are used to monitor and guide learners' progress (Alderson, Brunfaut, & Harding, 2015; Lee, 2015). There is also renewed interest in the washback effects of testing (Cheng, Watanabe, & Curtis, 2004), namely the way in which testing shapes learning through test preparation. Furthermore, self-assessment has been a crucial component of the Common European Framework of Reference (CEFR) (Little, 2005) and has gradually gained recognition and acceptance within Japan (Tono, 2013).

Thus, within the pedagogical context of the current study, language teachers in universities today are faced with the challenges of preparing learners for high-stakes assessments along with the challenge of incorporating assessments to monitor learner progress and provide learners with feedback that can benefit their future language learning. More research is needed to understand how or if existing assessment tools can be used efficiently and effectively for educational and pedagogical purposes. Therefore, the goal of the current study is

to conduct an exploratory study that focuses on the use of existing assessments for speaking within a Japanese university EFL context.

The purpose of the study was to explore the following research questions: (1) How do Japanese EFL learners perform on different types of oral proficiency assessments? and (2) What are Japanese EFL learners' perceptions of different types of oral proficiency assessments? The types of oral proficiency assessments used in the current study were a face-to-face interview test based on the IELTS model, a group semi-direct integrated speaking test based on the TOEFL iBT model, and a self-assessment sheet adapted from the Swiss European Language Portfolio.

Methods

Participants

Participants were 26 EFL learners enrolled at a women's university in Tokyo, Japan. They were English language majors in their third or fourth year of university studies with low intermediate to intermediate proficiency in English. Their background in studying English was typical of Japanese university students, with at least 6 years of formal English classes prior to entering university that mostly emphasized reading and writing skills. However, after beginning studies in the English department in their second year of university, they experienced a curriculum with content-based English classes, including lectures in English and student presentations, and structured instruction in academic reading and writing.

Four interviewers participated in the study. In order to preserve the naturalistic aspect of the setting, English instructors served as the interviewers. Two interviewers were native speakers of English, one was an experienced language instructor and one was a novice language

instructor. The other two interviewers were the first two authors of the current study. They are both bilingual speakers of English and Japanese and experienced language instructors at the university level. All of the interviewers had previous experience in oral test administration.

Materials and Procedure

Oral proficiency tests. Two types of oral proficiency tests were used for the study, a face-to-face direct interview test following the IELTS model, and a group test that followed the TOEFL model.

IELTS-type test. The IELTS-type test consisted of three parts and lasted for 15-20 minutes. In a one-on-one interview setting, learners were first asked questions about everyday topics related to home life, work, or study (Part 1). Then, learners were given a task card with a prompt. After one-minute of preparation time, learners spoke for 1 to 2 minutes in response to the prompt (Part 2). Finally, learners were asked further more abstract questions on the same topic (Part 3). The questions and task prompts used for the current study were taken from previous examinations published by University of Cambridge ESOL (Cambridge ESOL, 2011).

TOEFL-type test. The TOEFL-type test consisted of six questions and also lasted for 15 to 20 minutes. The questions were presented in a group setting, and learners recorded their responses individually on hand-held digital voice recorders. Questions 1 and 2 were independent speaking questions about familiar topics and students answered after 15 seconds of preparation time. Questions 3 to 6 were integrated speaking questions, that required learners to first listen to a dialogue or lecture and/or to read a short paragraph and then synthesize the source content in their responses after 20-30 seconds of preparation time.

Self-assessment. The self-assessment checklist was based on the Common European Framework of Reference (CEFR) level checklist in the Swiss version of the European Language Portfolio (Council of Europe, 2011). Learners were presented with 30 “can-do” statements describing various English actions such as “I can explain my hobby, club activities, or part-time job.” Learners were asked to evaluate whether they can do what is described in the statement on a three point scale: “can do it very easily”, “can do it”, “I don’t know”. There were six statements related to levels A1~B2, and four statements for level C1 (most advanced).

Questionnaires. Questionnaires were given to the learners immediately after the IELTS-type test (Questionnaire 1) and TOEFL-type test (Questionnaire 2), asking them about their experience on each of the tests. The questionnaires (see Appendix) asked learners to report on what they were paying attention to during the test by filling in a chart. The chart included five categories, pronunciation, vocabulary, grammar, and content. Learners filled in the chart with symbols that indicated how much attention they thought they had paid to each aspect of language. In addition, learners were also asked two open-ended questions about what was difficult about the test, and how they should prepare for the test before taking it again.

Procedure. Each learner participated in the IELTS-type test and TOEFL-type test and completed a questionnaire after each test. All tests were audio-recorded. The IELTS test was also video-recorded.

Analysis

The IELTS-type tests were rated by the four interviewers immediately after each interview based on the official published IELTS holistic rating criteria rubric. Each learner was rated on a scale from 0

to 9. Scores at half intervals (ex. 3.5, 4.5, and 5.5) were also allowed. The interviews conducted by the two native speakers (44% of the data) were also rated by a second rater. The first two authors of the study, who were also interviewers for the IELTS test, conducted the second ratings. Raters agreed perfectly or within half a band in 83% of the cases. The researchers' ratings were used in the analysis.

The TOEFL tests were also rated by the first two researchers. A sub-set of the data was rated by both of the raters. Simple percentage agreement for the holistic scores subset was 85%. In order to simulate oral proficiency testing in a university context, rating was conducted by English language instructors with experience in the university EFL context.

Results

Learners' Performance

The first research question focused on learners' performance on the two types of speaking tests and self assessment. Ratings for learners' performance on the IELTS-type test ($n=26$) ranged from 3 to 6, with a mean score of 4.44 ($SD=0.766$). Scores for learners' performance on the TOEFL-type test ranged from 4 to 20, with a mean score of 12.19 ($SD=4.5$). Finally, learners' self-assessment of their oral proficiency according to the CEFR levels ranged from A1 to B2 ($n=17$), with the largest number of learners responding to B1.

Pearson correlation coefficients were calculated to explore the relationship among learners scores for the two types of speaking tests and self-assessment levels. Results indicate a significant positive correlation between learners' scores on the two types of oral proficiency

tests ($r=.40$, $p<.05$). Furthermore, results also indicated negative (non-significant) correlations between learners' CEFR self-assessment levels and their oral proficiency test scores as is shown in Table 1.

Table 1. Correlation coefficients for test scores and self-assessment.

	IELTS	TOEFL	CEFR
IELTS	---	.40*	-.47
TOEFL		---	-.42
CEFR			---

Learners' Perceptions

The second research question focused on learners' perceptions about the two types of oral proficiency tests. Learners responses to Questionnaire 1, completed immediately after the IELTS-type test showed that on a scale of 0 – 3 (no attention to very much attention), learners paid most attention to content, then grammar and vocabulary, and least to pronunciation. The results for Questionnaire 2, completed immediately after TOEFL-type test, were very similar, as is shown in Table 2. Here too, learners paid most attention to content, then grammar and vocabulary, and least to pronunciation.

Table 2. Mean scores for aspects of language learners reported paying attention to during the oral proficiency test.

	Pronunciation	Vocabulary	Grammar	Content
IELTS-type	0.8	2.1	2.1	2.5
TOEFL-type	0.9	1.9	1.7	2.5

*Scale: 0 ~ 3

Questionnaire responses also indicated what learners considered to

be difficult about each of the tests. For Questionnaire 1, which focused on the IELTS-type test, 14 out of the 24 learners who responded to the questionnaire reported that content was the most difficult aspect of the test. Examples of learners' comments included, "I couldn't organize my thoughts," "I was not sure what I should say," and "I couldn't think of anything to say." In addition, 11 of the 24 learners reported that they had difficulties with using English. For example, learners made comments such as the following: "Even though I knew in Japanese what I wanted to say, I couldn't put my ideas in English because of lack of vocabulary knowledge," and "I didn't know how to say what I wanted to say."

As for Questionnaire 2, which focused on the TOEFL-type test, 14 of the 27 learners who completed the questionnaire reported that they had trouble with listening comprehension. In other words, these learners had difficulty understanding the input passage or conversation. Furthermore, 9 learners mentioned that the time element made the test difficult. Only 3 learners mentioned difficulty with content.

Finally, the questionnaires asked how learners thought they should study to prepare to take the same test again. After taking the IELTS-type test, 13 of the 24 learners answered that they needed some kind of output practice. For instance, one learners wrote "I need to get used to speaking in English," or "I should prepare so that I can talk about daily life in English" In addition, 10 of the 24 learners answered that they need to increase their English vocabulary knowledge. In contrast, after taking the TOEFL-type test, 16 of the 27 learners responded that they needed to improve their listening ability, and 11 of the 27 learners responded that they needed to increase their vocabulary.

Summary of Results

The results of the current study can be summarized as follows. First, with respect to learners' performance on the two types of oral proficiency tests, (1) a significant positive correlation was found between learners' performance on the face-to-face IELTS-type interview test and semi-direct TOEFL group test results, and (2) a negative correlation (non-significant) was found between learners' performance on the oral proficiency tests and self-assessment. Next, with respect to learners' perceptions, (1) for both types of tests, learners report focusing more on content than grammar and vocabulary and much less on pronunciation, (2) for the IELTS-type test, learners mainly reported difficulty with decisions on what to say and with English language use, and suggested a need to study vocabulary and increase output practice before taking the test again. (3) In contrast, for the TOEFL-type test, learners mainly reported difficulty with listening comprehension and time pressure, and suggested a need to improve their listening ability and also increase their vocabulary.

Discussion and Pedagogical Implications

Assessment of oral proficiency

The results of the study showed a significant positive correlation between the two types of oral proficiency tests, indicating that they both may serve a similar function when used to measure general oral proficiency within an educational context.

However, a negative correlation was found between the oral proficiency tests and the learners' self-assessment ratings. In other words, learners with higher test scores tended to have lower self-

assessment ratings, and learners with lower test scores tended to have higher self-assessment ratings. This mismatch between self-assessment and test scores is not surprising given the findings by Babaii, Taghaddomi, and Pashmforoosh (2015) in a recent study of speaking self-assessment. In a study of 29 university EFL students in Iran, Babaii et al. found significant statistical differences between teachers' and students' self-assessments of an IELTS-type speaking test. The mismatch between teachers' and students' ratings decreased after practice. The results of the current study should be interpreted with care because the self-assessment was based on can-do statements and the test scores were based on holistic rating criteria. Still, a negative correlation indicates that higher proficiency learners may tend to be more severe in their self-assessments whereas lower proficiency learners may tend to be more lenient in their self-assessments. Such results suggest that some sort of awareness raising or assessment literacy training may be necessary when incorporating self-assessment into the curriculum.

Learners' Perceptions

The results of the current study indicated that the two different types of oral proficiency assessments provided learners with different challenges. With regard to the IELTS-type test, learners reported difficulty with expressing their ideas and with deciding what to say. In contrast, learners reported difficulty with listening comprehension and time management with respect to the TOEFL-type test. Such differences in perceptions have implications for washback effects of the test. Learners reported that they would prepare for the IELTS-type test by studying vocabulary and practicing output, whereas learners reported that they would prepare for the TOEFL-type test also by

studying vocabulary and improving their listening comprehension. When using oral proficiency tests, not just for gate-keeping or placement purposes but for monitoring learner progress, washback effects may be an important consideration. In the current educational context, it could be said that the IELTS-type test would have a more direct impact on learners' subsequent efforts to improve their speaking skills.

However, it is striking that learners overwhelmingly reported for both types of tests that vocabulary knowledge was a major barrier to their success on the tests. Such results suggest that in either case, test washback effects would include an increased emphasis on vocabulary learning. Still, it is also important to note that in most cases learners' comments about how they would study for subsequent tests were broad statements that did not describe concretely how they would study. Such results indicate that the learners in the current study may need guidance in test preparation and that positive washback effects of the test may be limited without such support.

Finally, learners' comments showed that content also proved challenging for a number of learners. This is a point that should be highlighted in that it is an issue that is not only related to English language learning but more generally to learners' verbal communication skills.

Limitations and Conclusions

The current study explored Japanese university EFL learners' performance on two types of oral proficiency tests and self-assessments. The findings of the study indicate that both the IELTS-type and TOEFL-type may elicit similar results when used within the Japanese EFL

learners in a university setting. However, self-assessments should be introduced carefully and should be accompanied with support in order to elicit accurate results. More research is needed to investigate the content and extent of guidance learners may need to become competent in self-assessment.

When considered from a pedagogical perspective, the findings of the current study suggest that different washback effects may be expected depending on the type of test. Although both test types led learners to focus on their lack of vocabulary knowledge, the IELTS-type test was associated with a need for output practice, whereas the TOEFL-type test was associated with a need for listening comprehension practice. General verbal communication skills may also be key in preparing for the IELTS-type test. The findings also suggest that for positive test washback to occur, learners may need concrete support in planning their further studies. Further research is needed to investigate the pedagogical effects that oral proficiency testing can have on learning processes. It should be noted that the findings of the study were based on a limited data set in a particular instructional context. Therefore any generalizations must be made with care.

Acknowledgements

This research study was supported by MEXT Grant-in-Aid for Scientific Research (C) Grant Number 26370738 (Japan) . The author would like to thank the interviewers, raters, and research assistants for their help with the project.

References

- Alderson, C., Brunfaut, T., & Harding, L. (2015) . Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36, 236–260.
- Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2015) . Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, 32, 1–27.
- Cambridge ESOL (2011) . *Cambridge IELTS 8 Official Students' Book with Answers: Official Examination Papers from University of Cambridge ESOL examinations (IELTS Practice Tests)*. Cambridge University Press.
- Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in Language Testing: Research Contexts and Methods*. New York: Routledge.
- Council of Europe (2011). *Swiss version of the European Language Portfolio*. Retrieved from http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Key_reference/checklist_EN.pdf
- Educational Testing Service (2012). *Official guide to the TOEFL test with CD-Rom, Fourth Edition (Official guide to the TOEFL iBT)*. McGraw-Hill.
- Lee, Y. (2015). Diagnosing diagnostic language assessment. *Language Testing*, 32, 299–316.
- Little, D. (2005). The Common European Framework and European Language Portfolio: Involving learners and their judgements in the assessment process. *Language Testing* 22, 321–336.
- Tono, Y. (2013). *CAN-DO List Sakusei/ Katsuyou Eigotoutatsudo-mokuhyou CEFR-J Guidebook* [Guidebook for English Achievement Goals

The use of oral proficiency tests in the Japanese EFL context: Learners' perceptions
 CEFR-J: Creating and Utilizing CAN-DO Lists]. Tokyo: Taishukan.
 (投野由紀夫『CAN-DO リスト作成・活用 英語到達目標CEFR-J
 ガイドブック』)

Appendix.

Questionnaire

1. テストの時, どのようなことに気をとられていましたか? (Part 1, Part 2, Part 3それぞれについて◎, ○, △, ×などで教えてください)

◎=とても気をとられていた, ○=気をとられていた, △少し気をとられていた, ×=ほとんど考えていなかった

	Part 1	Part 2	Part 3
発音			
語彙			
文法			
内容 (何を話したら良いか)			
その他:具体的に			

2. 特にどのような点が難しかったですか?
3. もしまたこのようなテストを受けるとしたら, どのような準備をしたらよいと思いますか?