

Results of Pilot Tests that Measure Competencies of English Teachers in Japanese Elementary Schools

Mai Matsunaga

Abstract The development and results of the pilot English proficiency and teaching skills tests for English teachers in Japanese elementary schools are described in this paper. In order to measure current levels of oral proficiency and teaching skills of elementary school teachers, a listening test and an interview test in English were developed. The listening test consisted of questions adapted from a listening section of the EIKEN test grade 3 (Obunsha, 2007). The interview test consisted of two parts: the first part tested oral skills, especially speaking skills; and the second part tested teaching skills. Seven in-service elementary school teachers took the pilot tests in fall, 2008. To sum up the results of the pilot tests, the participants tended to achieve high levels in the listening and conversation sections while they tended to vary in their skills in answering questions on a paragraph, and teaching English. Based on the results and analysis of the pilot tests, a revised version will be developed for the second pilot tests.

Keywords: pilot tests, competencies, elementary school English teachers

小学校における日本人英語指導者を対象とした資質測定テストの結果

松 永 舞

要旨 この論文では、小学校における日本人英語指導者を対象として作成、実施した資質測定パイロットテストの結果を報告している。主にオーラルスキルと英語による指導力を測定するために作成されたパイロットテストは、リスニングテストとインタビューテストの2種類から構成されている。リスニングテストは実用英語技能検定試験（英検）3級の問題を使用する（Obunsha, 2007）。インタビューテストはスピーキング能力を測定するパート、指導力を測定するパートの2部構成である。完成したパイロットテストは2008年秋に7名の現職小学校教師を対象に実施された。結果をまとめると、参加者全体的にリスニング力、会話力は比較的高い一方で、パラグラフに関する質問への英語での応答、英語で指導する部分では個人差が目立った。今後は、今回のパイロットテストの結果分析をもとにテストを改善し、2回目のパイロットテストを実施する予定である。

キーワード: パイロットテスト、資質、小学校英語指導者

Purpose of Research

In March 2008, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) introduced a new course of study for the elementary school level, which included information about all classes offered in Japanese elementary schools. This course of study will be put into effect at the elementary school level in April, 2011. Within this new course of study, all fifth and sixth graders will be required to have a foreign language class, i.e., English class, once a week. There are three basic objectives to the new course of study: (a) to foster understanding of languages and cultures, (b) to promote active participation in communication, and (c) to develop basic communication skills (MEXT, 2008a, 2008b).

While English activities will be required at elementary schools, the issue over quality of teaching has been one of the obstacles elementary schools have faced. In other words, the level of English teachers in terms of their English ability and teaching skills has been at the center of discussion among researchers (e.g., Butler, 2005; Higuchi, Kanamori, & Kunikata, 2005). The importance of this issue over the quality of English teaching at Japanese elementary schools has motivated the author of this paper to conduct a 4-step study in an attempt to improve competence of elementary school English teachers. And the four steps are: (a) clarifying standards of competencies of Japanese elementary English teachers through organizing previous research results, (b) organizing information about examinations on these competencies of English teachers developed in Japan and other countries, (c) developing and implementing pilot examinations that evaluate competencies that Japanese elementary school English teachers should acquire, and (d) eventually, revising the examinations based on an analysis of the results and feedback from pilot test-takers.

Matsunaga (2008a), as a first step in the 4-step study, investigated levels of competencies Japanese elementary school English teachers need to obtain, through examining the current situation of English education at public elementary schools, and organizing previous research results. Following the first step, Matsunaga (2008b), as a second step, examined tests developed in Japan and other countries that were designed to measure the competencies of foreign language teachers. Moreover, Matsunaga (2008b, 2009a), as a third step, introduced an outline of measuring instruments she planned to create that would evaluate the competencies of English teachers in Japanese elementary schools. She also proposed a rubric of competence levels, which would be employed in the evaluation process. Furthermore, in Matsunaga (2009b), also

as part of the third step, pilot tests that the author completed based on the outline were introduced. Then, in this paper, she will report the results of these pilot tests administered to seven in-service elementary school teachers in fall, 2008.

Pilot Tests

Test Rationale

Using two measuring instruments, a listening test and an individual interview in English, current levels of English language ability and teaching skills of in-service Japanese English teachers at the elementary school level will be examined. Participants, in-service teachers, are expected to vary in their background of relevant knowledge or experience such as years of teaching English, experience abroad, or experience in attending teacher training sessions. Through the use of these measuring instruments, the participants will be able to recognize their current level of English and teaching skills, and use the results for further improvements. The results will be beneficial not only for the participants, but also for researchers, teacher trainers, and/or government officials who are in charge of elementary school English education because the results can be used as sources to gain an understanding of the current level of in-service teachers, to develop plans to improve the situation, and/or to organize appropriate training sessions.

Content Selection

A job analysis of elementary school English teachers was conducted, using an inventory the Language Testing Research Center (LTRC) at the University of Melbourne developed based on Ellis's (1984) inventory of teacher tasks (Elder, 1994). The LTRC's inventory included three major categories of teacher tasks: (a) interactions involving pedagogic goals which included medium-oriented interactions such as modeling the target language, message-oriented interactions such as explaining processes, and activity-oriented interactions such as giving instructions for a game; (b) interactions involving framework goals such as disciplining students; and (c) extra-classroom use of target language which included preparing the lesson, interacting with members of the school community such as talking to individual students about their progress or difficulties, and professional development such as reading professional journals. Based on the LTRC's inventory, a list of possible test or interview items for this study was created by the author. Then, the list was examined by two

groups in order to confirm content validity: (a) two university teachers who do research on elementary school English education, and also had teaching experience at the elementary school level, and (b) two in-service elementary school English teachers. Through this process, the items were refined to reflect what was expected of Japanese elementary school English teachers.

Development of Specifications

Test content and format

In order to measure participants' oral proficiency and teaching skills, a listening test and an interview in English were conducted. The listening test consisted of questions adapted from a listening section of the EIKEN test grade 3 (Obunsha, 2007 [see Matsunaga, 2008b for more detail]), and was given to all participants in the same room. The interview consisted of two parts: the first part tested oral skills, especially speaking skills, and the second part tested teaching skills. The interview was given individually by an interviewer and a rater for about 15 minutes. The first part included items such as a warm-up, everyday conversation, and reading a paragraph-long story aloud and answering questions about the story. This first part was meant to test content and manner of oral English. In other words, this part examined not only overall effectiveness, but also accuracy of sentence mastery and vocabulary, and control of fluency and pronunciation of oral production (e.g., Council of Europe, 2001; Harcourt Assessment, 2007). The second part included items such as giving instructions of a game, or acting out a dialogue with an interviewer. This part examined overall task fulfillment and completeness, recognition of student level, use of instructive language, and fluency (SEAMEO-RELC, NLLIA LTRC, & IKIP, 1997). The outline of the interview test is shown in Table 1 (see Matsunaga, 2009b for a complete interview test).

Table 1. *The outline of the interview test (Matsunaga, 2009b)*

Sections	Contents	Details	Time
Speaking skills			
I	warm-up conversation	greetings, Q & A	1 minute
II	reading aloud, Q & A	reading a short paragraph aloud, Q & A about the paragraph	1 minute & 30 seconds
Teaching skills			
I	giving instructions for a game		5 minutes
II	modeling a dialogue		3 minutes

Competence levels and rubrics

Listening skills, speaking skills, and teaching skills were separately assessed using different criteria. First, the listening skills were assessed based on a numerical score of correct responses to the total test items. Four levels were employed for the listening score: (a) level one, less than 50% of the total score; (b) level two, 50% to 59%; (c) level three, 60% to 69%; and (d) level four, more than 70%, with level three being set as a satisfactory level. Second, speaking skills were assessed based on the following six criteria: (a) overall effectiveness, (b) vocabulary, (c) grammar, (d) fluency, (e) coherence, and (f) pronunciation. Then, a holistic aspect, i.e., (a), and analytic aspects, i.e., (b), (c), (d), (e), and (f), of the spoken language use of each task were added and the average score was calculated as a speaking score. Four levels were employed for the speaking score: (a) level one, less than 60%; (b) level two, 60% to 69%; (c) level three, 70% to 79%; and (d) level four, more than 80%, with level three being set as a satisfactory level. Finally, teaching skills were assessed based on the following four criteria: (a) overall task fulfillment, (b) recognition of student level, (c) instructional language, and (d) fluency. Then, a holistic aspect, i.e., (a), and analytic aspects, i.e., (b), (c), and (d), of the teaching skills of each task were added and the average score was calculated as a teaching skills score. Four levels were employed for the teaching skills score: (a) level one, less than 60%; (b) level two, 60% to 69%; (c) level three, 70% to 79%; and (d) level four, more than 80%, with level three being set as a satisfactory level.

Moreover, the author developed rubrics for the three skills, i.e., listening, speaking, and teaching skills, which were used in the evaluation process. In the development of the rubrics, the Common European Framework (CEF) (Council of Europe, 2001), the Canadian Language Benchmarks (CLB) (Smith, 2002), and the SEAMEO-RELC, NLLIA LTRC, & IKIPS criteria (1997) were used as references. In other words, competence levels in this study correspond to those of these three references. The CEF is suitable as a universal standard since it has been used worldwide. The CLB is a Canadian standard used for describing, measuring and recognizing the second language proficiency of adult immigrants or prospective immigrants for living and working in Canada, and it provides a descriptive scale of communicative proficiency in English as a second language, covering four skill areas (Center for CLB, retrieved in 2007). Since the CLB gives concrete descriptions of standards or criteria that match the CEF scales of overall proficiency, it is used in the study in addition to the CEF. The SEAMEO-RELC, NLLIA LTRC, & IKIPS criteria is used as the corresponding

reference for teaching skills in the study (see Matsunaga, 2009a, 2009b for more information on the CEF, CLB, & SEAMEO-RELC, NLLIA LTRC, & IKIPS criteria, and the original rubrics).

Selection and Training of Raters

One interviewer and rater were selected for the administering of the pilot version interview. They were Japanese university English teachers who have a high level of oral proficiency in English. They took training sessions beforehand on the rationale and specifications of the instruments, focusing especially on the rating scales for the interview. Moreover, they had a chance to actually rate sample interviews for practice. In the actual administering of the interviews, each interview was conducted by both an interviewer and a rater, and was video-recorded. After the initial interview, the original interviewer and rater reviewed the video-recording and re-rated it together. If they disagreed on their rating, they discussed it until they agreed on the same rating.

Results of the Pilot Tests

It is not possible to statistically generalize the results of the pilot tests for all elementary school English teachers in Japan because of a limited number of participants (a total of seven participants). However, the author still believes that it is worthwhile to statistically process the results of the pilot tests in order to describe the overall tendency among the participants, and possibly of elementary school teachers in general.

Method

Participants and the recruitment procedure

A teacher trainer in Higashi-Osaka city, Osaka, helped to recruit participants for the pilot tests. He recruited seven in-service elementary school teachers in the district. The participants were all regular homeroom teachers of fifth or sixth graders from various schools in the district, and were teaching English as homeroom teachers at the time of testing. Five of the teachers were female and two of them were male. Their age and English teaching experience varied. Since all the participants were leaders of English education at their schools, the level of their English proficiency or teaching skills was expected to be higher than that of the average elementary school teachers.

Procedures

The pilot tests were administered after a seminar by the Higashi-Osaka Board of Education in fall, 2008. The listening test was administered first for about 40 minutes, and then each participant took an interview test with an interviewer and a rater for about 15 minutes. All the interviews were administered by the same interviewer and rater. In addition, all interviews were video-recorded with permission by the participants.

Results and discussion

Listening test

As Table 2 shows, the average total score of the listening test surpassed a passing score, which was set at 18 points (60%). In addition, six out of the seven participants gained the passing score and five of them obtained the score necessary for level four (the highest level), i.e., more than 21 points (over 70%). These results revealed that the majority of the participants had the level of listening skills that was considered as a requirement to be a successful English teacher at the elementary school level in this study. Considering the fact that the participants or general elementary school teachers usually have a university degree, the listening test employed in this study, which tests junior high school level English, may have been rather easy for the participants. However, since all the participants were leaders of English education at their schools, the level of their listening skills may have been higher than that of average elementary school teachers.

Table 2. Means and Standard Deviations for Listening Test Scores

	<i>M</i>	<i>SD</i>
Section 1	8.43	1.72
Section 2	7.71	2.36
Section 3	8.57	1.40
Total	24.71	4.92

Note. Types of questions were the following: section 1, questions and responses; section 2, short conversations; and section 3, passages. Each section had 10 questions, and each question had 1 point, making the highest possible total score 30 points. There were seven participants for the test.

Interview test, speaking skills test

The speaking skills test consisted of two sections: a conversation, and reading a paragraph aloud and answering questions. The former included two parts: greetings, and questions and responses. The latter also included two parts: reading a short paragraph aloud, and answering three questions about the paragraph. Looking at the average total score of speaking skills in Table 3, the average total score surpassed a passing score, which was set at 70 points (level three). Four out of the seven participants gained the passing score and three of them obtained the score necessary for level four (the highest level), i.e., more than 80 points (over 80%); on the other hand, one participant was assessed at level two, i.e., less than 70 points (below 70%) and two participants were assessed at level one, i.e., less than 60 points (below 60%). Since the participants in this pilot test were thought to have a higher level of English proficiency than that of average elementary school teachers, it was assumed that average elementary school teachers would not reach a level which was as high as that of these participants.

Looking at the results of the conversation section in Table 3, the average total score surpassed a passing score, which was set at 70 points (level three). Moreover, the average scores of both parts in the conversation section, i.e., greetings, and questions and responses, surpassed the passing score, which was set at 70 points (level three). Furthermore, six out of the seven participants gained the passing score and five of them obtained the score necessary for level four (the highest level), i.e., more than 80 points (over 80%). These results revealed that the majority of the participants had good control of everyday English conversation.

Looking at the results of the section, reading a paragraph aloud and answering questions, in Table 3, the average total score fell short of a passing score, which was set at 70 points (level three). Four out of the seven participants gained the passing score and three of them obtained the score necessary for level four (the highest level), i.e., more than 80 points (over 80%); on the other hand, three participants were assessed at level one, i.e., less than 60 points (below 60%). As Table 4 shows, this discrepancy in scores of the participants was due to the fact that those who achieved higher levels gained high scores in the part, answering questions; on the contrary, those who achieved lower levels gained low scores in the same part. There were mainly three reasons for their low scores in the part, answering questions: (a) they did not understand the questions (i.e., problems with listening skills), (b) they did not find the answers in

the paragraph (i.e., problems with reading skills), or (c) they could not express their answers in English (i.e., problems with speaking skills). Since this part required integrated skills including all three of these skills, it was assumed that some of the participants found the part beyond their ability. With the development of students' communication skills being one of the objectives in English activities at the elementary school level, the development of teachers' communication skills, especially their listening and speaking skills, will be crucial for successful implementation of the programs. On the other hand, the average score for the part, reading a paragraph aloud, surpassed the passing score, which was set at 70 points (level three); all the participants reached the highest level, level four, in this part. This result revealed that all the participants had good control of reading a short paragraph aloud with appropriate pronunciation, intonation, and fluency to be a model for their students. Considering the fact that the paragraph in the test was short and set at the level of junior high school English, it may have been rather easy for the participants or elementary school teachers in general.

Table 3. Means and Standard Deviations for Speaking Skills Test Scores

	<i>M</i>	<i>SD</i>
Speaking skills Total	70.29	10.84
Conversation Total	79.86	10.11
Greetings	79.86	0.38
QAs	73.86	14.97
Reading aloud & QAs Total	64.57	17.50
Reading aloud	80.0	0.0
QAs	59.43	23.34

Note. The speaking skills test consisted of two sections: conversation, and reading a paragraph aloud and answering questions. The former included greetings, and questions and answers. The latter included reading a short paragraph aloud, and answering questions about the paragraph. The highest possible score of each category was 80 points. There were seven participants for the test.

Interview test, teaching skills test

The teaching skills test consisted of two sections: giving instructions for a game, and modeling a dialogue. Looking at the total score of teaching skills in Table 4, the average score fell short of a passing score, which was set at 70 points (level three). This rather low average can be attributed to a large gap in scores between the higher level participants and the lower level participants, with the latter group

scoring extremely low. Four out of the seven participants gained the passing score and two of them obtained the score necessary for level four (the highest level), i.e., more than 80 points (over 80%); on the other hand, one participant was assessed at level two, i.e., less than 70 points (below 70%), and two participants were assessed at level one, i.e., less than 60 points (below 60%). Since the participants in this pilot test were thought to have a higher level of teaching skills than those of average elementary school teachers, it was assumed that average elementary school teachers would not reach a level which was as high as that of these participants.

Looking at the results of the section, giving instructions for a game, in Table 4, the average score fell short of a passing score, which was set at 70 points (level three). This rather low average was due to a discrepancy in scores of the participants. Five of the seven participants gained the passing score and two of them obtained the score necessary for level four (the highest level), i.e., more than 80 points (over 80%); on the other hand, two of the participants were assessed at level one, i.e., less than 60 points (below 60%).

Looking at the results of the section, modeling a dialogue, in Table 4, the average score fell short of a passing score, which was set at 70 points (level three). This rather low average was due to a discrepancy in scores of the participants. Four of the seven participants gained the passing score with two of them obtaining the score necessary for level four (the highest level), i.e., more than 80 points (over 80%); on the other hand, two of the participants were assessed at level two, i.e., less than 70 points (below 70%), and one participant was assessed at level one, i.e., less than 60 points (below 60%). In addition, those who achieved higher levels in the section, giving instructions for a game, also tended to achieve higher levels in the section, modeling a dialogue.

Through the administration of the pilot teaching skills test, three problems with the test became evident. First, although the test was originally meant to evaluate the participants' teaching skills, it evaluated more of their English proficiency than their teaching skills. In other words, the test tended to evaluate how fluently they spoke English rather than how well they conveyed necessary information to their students. Second, the test cards that the participants received for each section included too much information to process within the time they had been given. This also led the participants only to read out the instructions or dialogue in English, not paying too much attention to the fact that they were supposed to be talking to their students.

Finally, the section, modeling a dialogue, did not reflect the participants' actual teaching at their schools. This means that most participants did not seem to understand what they were asked to do with the dialogue although the instructions for the test were explained on the card. Keeping these problems in mind, the test should be revised to more effectively assess teaching skills, and reflect actual teaching situations.

Table 4. Means and Standard Deviations for Teaching Skills Test Scores

	<i>M</i>	<i>SD</i>
Teaching skills Total	67.0	17.04
Game instructions	64.14	29.79
Modeling a dialogue	69.29	12.05

Note. The teaching skills test consisted of two sections: giving instructions for a game, and modeling a dialogue. The highest possible score of each section was 80 points. There were seven participants for the test.

Final Remarks

The author developed pilot instruments that evaluated the competence of English teachers in Japanese elementary schools. These instruments focused on oral English ability and teaching skills, and therefore, they included a listening test that evaluated listening skills, and an interview test in English that evaluated speaking and teaching skills. In the evaluation process, listening skills, speaking skills, and teaching skills were separately assessed using different criteria and level descriptions, which corresponded to the CEF, CLB, and SEAMEO-RELC, NLLIA LTRC, & IKIP scales. In this paper, the results of the pilot tests were mainly discussed.

Seven in-service elementary school teachers, who were the leaders of English education at their schools, took the tests in fall, 2008. Regarding the listening test, the average total score surpassed a passing score, which was set at 18 points (60%). In addition, six out of the seven participants gained the passing score and five of them obtained the score necessary for level four (the highest level), i.e., more than 21 points (over 70%). These results revealed that the majority of the participants had the level of listening skills that was considered as a requirement to be a successful English teacher at the elementary school level in this study. Regarding the speaking skills test in the interview test, the overall average speaking skills score surpassed a passing score, which was set at 70 points (level three). However, looking at the results of the section, reading a paragraph aloud and answering questions, the average of the total

score fell short of the passing score. This was mainly due to the fact that three participants achieved only low levels in the part, answering questions, and this fact lowered the average score of the whole section. Regarding the teaching skills test in the interview test, the average total score of teaching skills fell short of a passing score, which was set at 70 points (level three). In addition, the average score of each section in the teaching skills test, i.e., giving instructions for a game, or modeling a dialogue, fell short of the passing score. This rather low average was due to a discrepancy in teaching skills of the participants. Therefore, to sum up the results of the pilot tests, the participants tended to achieve high levels in the listening and conversation sections while they tended to vary in their skills in answering questions on a paragraph, and teaching English.

Even though this study had only a limited number of participants, and therefore, the data may not be statistically valid, the author believes that the data can show an overall tendency of elementary school teachers to an extent. Based on the results and analysis of the pilot tests, a revised version will be developed for the second pilot tests. Major revisions will have to be made in the teaching skills test in order to more appropriately reflect the level set as necessary for a successful English teacher at the elementary school level in the study. In other words, the level of the pilot teaching skills test was rather high, and the test contents did not necessarily reflect actual teaching situations in elementary schools. The author hopes that her study and measuring instruments will contribute to teacher development and further improvements in elementary school English education in Japan.

References

- Butler, Y. G. (2005). *English language education in Japanese elementary schools: Analyses and suggestions based on East Asian perspectives*. Tokyo: Sanseido.
- Center for Canadian Language Benchmarks. What are the Canadian language benchmarks? Retrieved September 14, 2007, from http://www.language.ca/display_page.asp?page_id=206
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Elder, C. (1994). Performance testing as benchmark for LOTE teacher education. *Melbourne Papers in Language Testing*, 3(1), 1-25.
- Ellis, R. (1984). *Classroom second language development*. New York: Prentice Hall.

- Harcourt Assessment. (2007). *Versant for English technical manual*. CA: Harcourt Assessment. Retrieved November 6, 2007, from <http://www.ordinate.com>
- Higuchi, T., Kanamori, T., & Kunikata, T. (Eds.). (2005). *Korekara no shogakkou eigokyoiku riron to jissen* [Future English education at elementary schools: Theories and practices]. Tokyo: Kenkyusha.
- Matsunaga, M. (2008a). Aspects relevant to English teacher competence in Japanese elementary schools. *Kinki University English Journal*, 1, 35-51.
- Matsunaga, M. (2008b). An outline of instruments to measure competence of Japanese English teachers at the elementary school level. *Kinki University English Journal*, 2, 1-18.
- Matsunaga, M. (2009a). A rubric of competence levels of Japanese English teachers in elementary schools. *Kinki University English Journal*, 3, 1-18.
- Matsunaga, M. (2009b). Tests to measure competencies of Japanese English teachers at the elementary school level. *Kinki University English Journal*, 4, 13-29.
- Ministry of Education, Culture, Sports, Science and Technology. (2008a). *Sho chu gakushusidoyoryo an* [Drafts of courses of study for elementary and junior high schools]. Retrieved February 15, 2008, from <http://www.mext.go.jp/>
- Ministry of Education, Culture, Sports, Science and Technology. (2008b). *Shogakko gakushu shido yoryo* [A course of study for elementary schools]. Tokyo; Tokyoshoseki.
- Obunsha. (2007). *Eiken sankyu zenmondaishu* [The EIKEN test in practical English proficiency grade 3]. Tokyo: Obunsha.
- SEAMEO-RELC, NLLIA LTRC, & IKIPS. (1997). *The English Proficiency Test for Indonesia: EPTI Manual*. Singapore: SEAMEO RELC.
- Smith, G. P. (2002). *Canadian language benchmarks 2000: Theoretical framework*, ON: Center for Canadian Language Benchmarks. Retrieved September 14, 2007, from <http://www.language.ca>