

Improving an Instrument that Measures Competencies of Japanese English Teachers in Elementary Schools

Mai Matsunaga

Abstract Based on the results and analysis of a first set of pilot tests, a second set of pilot English proficiency and teaching skills tests for English teachers in Japanese elementary schools was developed. In this paper, the results and examination of the second set of pilot tests are explained. 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 conducted. 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. Fifteen in-service elementary school teachers took the second set of pilot tests in 2009. To sum up the results of these pilot tests, the participants achieved high levels on the listening test while they tended to achieve only low levels on the interview test. Furthermore, examination of the contents of the second set of pilot tests revealed that the changes in the interview test helped to improve the quality of the test.

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改訂版資質診断テスト —結果と改訂内容の検証—

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、実施した第1回資質診断パイロットテストの結果と分析をもとに改訂された、第2回パイロットテストの結果と改訂内容の検証について報告している。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級の問題を使用した（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部構成であった。第2回パイロットテストは2009年に15名の現職小学校教員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テストの得点は高かったが、インタビューテストの得点は低かった。さらに、改訂内容の検証を行ったところ、インタビューテストの改訂内容がテストの質を高めたことが分かった。

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 an understanding of languages and cultures, (b) to promote active participation in communication, and (c) to develop basic communication skills (MEXT, 2008).

While English activities will be required in elementary schools, the issue over the 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 in Japanese elementary schools has motivated the author of this paper to conduct a 4-step study in an attempt to improve the competence of elementary school English teachers (ESET). And the four steps are: (a) clarifying standards of competencies of Japanese ESET 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 ESET should acquire, and (d) eventually, developing more reliable and valid examinations based on an analysis of the results and feedback from pilot test-takers. In this paper, as part of the third step, i.e., (c), the author will report the results of the second set of pilot tests administered to 15 in-service elementary school teachers in 2009 (see Matsunaga, 2010 for the results of the first set of pilot tests). She will also compare the second set of pilot tests with the first set of pilot tests in order to ascertain whether improvements were made in content between the two tests.

Method

Participants

A total of 15 in-service public elementary school teachers in Osaka and Hyogo, Japan participated in the study in 2009. These participants were all regular

homeroom teachers of different grade levels. Five of them were teaching English once a week as homeroom teachers at the time the test was given, but the other 10 had not had experience teaching English. Ten of them were female and five of them were male teachers, and the majority of them were in their twenties. In addition, those who were teaching English at the time of testing all had experience team-teaching with ALTs, and four of them had been abroad for more than a week. Unlike the participants in the first set of pilot tests who were the leaders of English teachers at their schools, all the participants in the second set of pilot tests were regular homeroom teachers. Therefore, the level of their English proficiency or teaching skills was expected to more accurately represent that of average elementary school teachers in Japan, compared with the participants in the first sets of pilot tests.

Measure

Test rationale

Using two measuring instruments, a listening test and an individual interview in English, current levels of oral and teaching skills of in-service Japanese English teachers at the elementary school level were examined. Participants were 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 instruments, the participants were able to recognize their current level of English and teaching skills, and use the results for further improvements. The results will also be beneficial for researchers, teacher trainers, and/or government officials who are in charge of elementary school English education, 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 ESET was conducted, using an inventory of teacher tasks, the Language Testing Research Center (LTRC) at the University of Melbourne developed based on Ellis' (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, and professional development such as reading professional journals. Based on the LTRC's inventory, a list of possible test items for this study was created by the author. Then, the list was examined by two groups: (a) two university teachers who do research on elementary school English education, and are also involved in the training of ESET, and (b) two in-service ESET. Through this process, the items were refined to reflect what was expected of Japanese ESET (see Matsunaga, 2009b for a complete interview test).

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 30 questions adapted from a listening section of the STEP test grade 3 (Obunsha, 2007). The interview consisted of two parts: the first part tested content and manner of oral English, and the second part tested teaching skills. The first part included items such as a conversation, and reading a paragraph-long story aloud and answering questions about the story. The second part included items such as giving instructions of a game, and acting out a dialogue with an interviewer. The outline of the interview test is shown in Table 1.

Table 1. *The outline of the interview test, second pilot test (Matsunaga)*

Sections	Contents	Details	Time
Speaking skills			
I	conversation	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		7 minutes
II	modeling a dialogue		5 minutes

Competence levels and rubrics

Regarding the satisfactory levels of English language ability and teaching skills in this study, the author suggests that, based on previous studies (e.g., Butler, 2005; Higuchi, Kanamori, & Kunikata, 2005), ESET should have a level of oral English

abilities with which they can listen to or speak junior high school graduate level English. Moreover, they should be able to use commonly used English expressions in teaching situations, and to communicate or team-teach with Assistant Language Teachers (ALT). Regarding teaching skills, the author suggests that teachers should have teaching skills that include two aspects: they should be able to utilize authorized textbooks such as English Notebook 1 and 2 (MEXT, 2009) along with their teacher's manuals; and they should be able to conduct activities mostly in English, with appropriate pronunciation and intonation which can be understood by native speakers of English who are accustomed to communicating with language learners. Based on these satisfactory levels in the study, 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 for listening and speaking skills, the Common European Framework of Reference (CEFR) (Council of Europe, 2001) was used as a reference. In other words, the competence levels of listening and speaking skills in this study correspond to those of the CEFR. The CEFR was chosen since it has been universally used as a reference for second or foreign language learning (e.g., Educational Testing Service, 2004). In addition, the rubrics for teaching skills employed the four categories introduced by the SEAMEO -RELC, NLLIA LTRC, & IKIPS (1997). However, the SEAMEO -RELC, NLLIA LTRC, & IKIPS did not reveal level descriptions, and therefore, the author created her own level descriptions based on their four categories (see Matsunaga, 2009a for more information on the author's original 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 (CEFR A1), less than 50% of the total score; (b) level two (CEFR Lower A2), 50% to 59%; (c) level three (CEFR Upper A2), 60% to 69%; and (d) level four (CEFR B1), 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 category, i.e., (a), and analytic categories, i.e., (b), (c), (d), (e), and (f), of each task were added and the average score was calculated for each section. Four levels were employed for the speaking score: (a) level one (CEFR A1), less than 60%; (b) level two (CEFR Lower A2), 60% to 69%; (c) level three (CEFR Upper A2), 70% to 79%;

and (d) level four (CEFR B1), 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 category, i.e., (a), and analytic categories, i.e., (b), (c), and (d), were added for each section. 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.

Procedure

Pilot tests

The listening test was administered first for approximately 30 minutes, and then each participant took the interview test. Each interview was conducted by an interviewer and a rater, and lasted for approximately 15 minutes. All the interviews were video-recorded with permission by the participants. After the initial interview, the interviewer and rater reviewed the video-recording and re-rated it together.

Examination of the second set of pilot tests

The contents of the second set of pilot tests were examined for improvements based on the following three sources: (a) the results and analyses of the first and second sets of pilot tests, (b) the video-recording of the interviews in the first and second sets of pilot tests, and (c) the results of a questionnaire completed by the participants in the second set of pilot tests (see the questionnaire in the Appendix).

Selection and Training of Raters

One interviewer and rater were selected for the administering of the 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.

Results and Discussion

It is not possible to statistically generalize the results of the second pilot tests to all ESET in Japan because of a limited number of participants in the study (a total of 15 participants). Despite this limitation, 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.

Descriptive Statistics

Listening test

As Table 2 shows, the average total score of the listening test exceeded a passing level, which was set at 18 points (60%). In addition, all the participants gained the passing score and 10 of them obtained the score necessary for level four. These results revealed that all 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.

Table 2. Means and standard deviations for listening test scores ($n=15$)

	<i>M</i>	<i>SD</i>
Section 1	8.47	1.19
Section 2	8.00	1.20
Section 3	8.80	1.86
Total	25.27	3.37

Note. Types of questions were the following: Section 1, question and response; Section 2, short conversations; and Section 3, passages. Each section had 10 questions, and each question was worth 1 point, making the highest possible total score 30 points.

Interview test: Speaking skills test

Looking at the average total score of speaking skills in Table 3, the average score did not reach a passing level, which was set at 70 points. Four out of the 15 participants reached the passing level, but nobody obtained the score necessary for level four. On the other hand, as the large SD implies, eight participants were assessed at level one. These results revealed that the majority of the participants did not exhibit adequate control in speaking skills.

Table 3. Means and standard deviations for speaking skills test scores ($n=15$)

	<i>M</i>	<i>SD</i>
Speaking skills total	56.20	17.13
Conversation total	51.33	26.52
Reading aloud & QAs total	59.44	14.58

Note. The highest possible score of each section was 100 points. The Speaking skills total consisted of 40% of the Conversation total, and 60% of the Reading aloud & QAs total.

Interview test: Teaching skills test

Looking at the average total score of teaching skills in Table 4, the average score did not reach a passing level, which was set at 70 points. Four out of 15 participants gained the passing score, but none of them obtained the score necessary for level four. On the other hand, as the large SD implies, nine participants were assessed at level one. These results revealed that the majority of the participants did not exhibit adequate control in teaching skills. In addition, the participants who gained higher scores in the section, giving instructions for a game, also tended to gain higher scores in the section, modeling a dialogue.

Table 4. Means and standard deviations for teaching skills test scores ($n=15$)

	<i>M</i>	<i>SD</i>
Teaching skills total	39.69	27.59
Game instructions	37.13	32.15
Modeling a dialogue	43.53	28.72

Note. The highest possible score of each section was 100 points. The Teaching skills total consisted of 60% of the Game instructions total, and 40% of the Modeling a dialogue total.

Examination of the Second Set of Pilot Tests

Listening test

According to the results of the questionnaire completed by the participants in the second set of pilot tests, all the participants evaluated the contents of the listening test appropriate to assess the listening skills required of an elementary school teacher. Examining the results of both the first and second sets of pilot tests, however, the passing level of the listening test, i.e., 60% (18 points), was too low and therefore, even teachers whose ability was not high enough to be a successful teacher at the elementary school level, could attain a passing score. In other words, the participants who barely reached the passing level did not have good ability in everyday English conversation or answering the questions about the paragraph, which required basic listening skills necessary for a successful teacher in elementary schools. Looking at the results, the passing level should be set at 80% (25 points) in the study, since the participants who reached this level tended to have good control of everyday English conversation or ability in answering the questions about the paragraph. Therefore, the rating scale for the listening test will have to be changed to the following: (a) level one, below 60%; (b) level two, 60 to 79%; (c) level three, 80 to 89%; and (d) level four, over 90%, with

level three being a satisfactory level.

Speaking test

According to the results of the questionnaire completed by the participants in the second set of pilot tests, all the participants evaluated the contents of the speaking test appropriate for assessing the speaking skills required of an elementary school teacher. Examining the results of the conversation section of both the first and second sets of pilot tests, the number of questions, i.e., two questions, was not sufficient for assessing the level of conversation skills, since some participants achieved level four on one question, but only level one on the other question, leading them to receive a low level in the section as a result. In order to more accurately assess conversation skills of participants, the number of questions should be increased to four or five in the future test.

The other section of the speaking test, reading a paragraph aloud and answering three questions about the paragraph, yielded a low average in both tests. Examining these results, some participants achieved level four on one or two questions about the paragraph, but only level one on another question, leading them to receive a low level in the part as a result. Therefore, the number of questions should be increased from three to four or five in the future test in order to more accurately assess oral fluency of participants. However, the author still believes that the passing level of both parts in the section was appropriate as a required level for a successful elementary school teacher since both parts were set at the level of junior high school English, which the author believed to be the minimum level for an elementary school teacher. In addition, according to the results of the questionnaire completed by the participants in the second set of pilot tests, all the participants evaluated the passing level of both parts in the section appropriate for an English teacher at the elementary school level.

Teaching skills test

Three problems with the first teaching skills test became evident, and they were: (a) the test tended to evaluate more of the participants' English proficiency than their teaching skills, (b) the test cards that the participants received for each section included too much information to process within the time they had been given, and (c) most participants did not seem to understand what they were asked to do with the

dialogue. In order to more effectively assess participants' teaching skills, the following changes were made for the second teaching skills test. For the section, giving instructions for a game, participants were asked to give the whole set of instructions in the first test, but were only asked to give part of the whole set of instructions in the second test. In addition, in order to reflect actual teaching situations, participants were required to memorize the instructions and act them out as if they were teaching their students. For the section, modeling a dialogue, in the first test, participants were asked to explain the main characters and situation of the dialogue first, and then act out the dialogue, pretending to be an elementary school student. In the revised version, they were asked to introduce the topic of a new dialogue, using picture cards, and then act out the dialogue as a homeroom teacher, interacting with an ALT (an interviewer). Moreover, the dialogue was shorter, focusing on introducing a target sentence for students to practice. Furthermore, for both sections, the following three changes were made: (a) contents of the test card were orally explained by the interviewer, while participants were looking at the card; (b) participants had more preparation time; and (c) participants were more encouraged to act out instructions or a dialogue as if to their own students.

Examining the results of the second set of pilot tests and video-taped interviews, the three problems with the first teaching skills test were minimized in the second set of pilot tests. Regarding the first problem of the test tending to evaluate more English proficiency than teaching skills, the participants seemed to concentrate on their teaching because they were clearly instructed to act out instructions or a dialogue as if to their own students, and they also had more preparation time. Considering the second problem of the test cards including too much information, the participants did not seem to have difficulty comprehending what was written on the cards in the second set of pilot test since the cards included less information and the interviewer orally explained the contents of the cards. Regarding the third problem of participants not understanding what they were expected to do in the section, modeling a dialogue, most participants did not seem to have difficulty comprehending what was expected of them. Moreover, according to the results of the questionnaire completed by the participants in the second set of pilot tests, they all evaluated both the contents and the passing level of the teaching skills test appropriate to assess elementary school teachers.

Conclusion

The author developed the revised version of the pilot instruments that evaluate the competence of Japanese ESET. These instruments focused on oral English ability and teaching skills, and therefore, they included a listening test, and an interview test in English. In the evaluation process, listening skills, speaking skills, and teaching skills were separately assessed using different criteria and level descriptions, which correspond to the CEFR, and the SEAMEO-RELC, NLLIA LTRC, & IKIP criteria. In this paper, the results and examination of the second set of pilot tests were discussed.

Fifteen in-service elementary school teachers took the pilot tests in 2009. Regarding the listening test, the average total score of the test (25.27) exceeded a passing level, which was set at 18 points (60%). The results revealed that all the participants had attained 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. While all the participants gained high scores on the listening test, they did not achieve a passing level (70%) on the speaking or teaching skills tests. These results revealed that the majority of the participants lacked oral fluency and teaching skills in English. Furthermore, the author compared the second set of pilot tests with the first set of pilot tests in order to ascertain whether improvements were made in content between the two tests, and concluded that the changes in the interview test helped to improve the quality of the test.

Even though this study had only a limited number of participants (N=15), and therefore, the data may not be statistically valid, the author believes that the data can show an overall tendency of elementary school teachers. Based on the results and examination of the second set of pilot tests, a more reliable and valid version will have to be developed for future use.

References

- Butler, Y. G. (2005). *English language education in Japanese elementary schools: Analyses and suggestions based on East Asian perspectives*. Tokyo: Sanseido.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Educational Testing Service. (2004). *Mapping TOEFL, TSE, TWE, and TOEIC on the Common European Framework executive summary*. NJ: ETS. Retrieved March 10, 2008, from <http://www.ets.org/>

- 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.
- 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. (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.
- Matsunaga, M. (2010). Results of pilot tests that measure competencies of English teachers in Japanese elementary schools. *Kinki University English Journal*, 5, 1-13.
- Ministry of Education, Culture, Sports, Science and Technology. (2008). *Shogakko gakushu shido yoryo* [A course of study for elementary schools]. Tokyo: Tokyoshoseki.
- Ministry of Education, Culture, Sports, Science and Technology. (2009). *Eigo noto 1 & 2* [English Notebook 1 & 2]. Tokyo: Kyoikushuppan.
- 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.

Appendix

Questionnaire on the Second Set of Pilot Tests

パイロットテストに関するアンケート

本日はお忙しい中、パイロットテストを受験いただきありがとうございます。以下のアンケートはテストを改善していく過程で参考にさせていただきたいと思います。お疲れのところ大変恐縮ですが、ご記入いただけると幸いです。

ご自身に関して

1. 年齢 (○で囲んでください) 20代 30代 40代 50代 60代
2. 小学校教師経験年数 () 年目
3. 小学校英語指導経験年数 () 年目
4. ALT と一緒に教えた経験 (○で囲んでください) ある ない
5. 資格試験の経験 (あれば) 英検 () 級、 TOEIC () 点
その他 ()
6. 海外留学経験の有無 (1週間以上あれば)
国 () 期間 ()

リスニングテストに関して (最適なものを一つ選んで○で囲んでください)

1. リスニング問題全体の難易度をどう感じましたか。
とても簡単 やや簡単 普通 やや難しい とても難しい
2. リスニング問題の中で最も難度の高いパートはどれでしたか。
第1部 (応答を選ぶ問題)
第2部 (対話問題)
第3部 (長めの英文問題)
3. ご自身のリスニングテストの結果は以下のどれだと想像されますか。
ほぼ満点 (9割以上)、 よくできた (8割以上)、 普通 (6割以上)
あまりできなかった (5割以上)、 ほとんどできなかった

4. このリスニングテストの合格点（6割以上）は、小学校英語指導者のリスニング力として適切だと思いますか。

適切

低すぎる（→理由： _____）

高すぎる（→理由： _____）

インタビューテストに関して（最適なものを一つ選んで○で囲んでください）

I. スピーキング能力をみるための質問に関して

1. 短いパラグラフの音読の難易度をどう感じましたか。

とても簡単 やや簡単 普通 やや難しい とても難しい

2. 音読後の質疑応答問題の難易度をどう感じましたか。

とても簡単 やや簡単 普通 やや難しい とても難しい

3. ご自身の音読・質疑応答の結果は以下のどれだと想像されますか。

ほぼ満点（9割以上）、 よくできた（8割以上）、 普通（7割以上）、
あまりできなかった（6割以上）、 ほとんどできなかった

4. この音読・質疑応答の合格点（7割以上）は、小学校英語指導者の力として適切だと思いますか。

適切

低すぎる（→理由： _____）

高すぎる（→理由： _____）

II. 指導能力をみるための質問に関して

1. 「ゲームのやり方の説明」の難易度をどう感じましたか。

とても簡単 やや簡単 普通 やや難しい とても難しい

その他ご意見・ご感想・ご質問等がございましたら以下にご記入ください。



ご協力ありがとうございます。
アンケートを封筒に入れてください。