

An Evaluation Method to Assess Sixth-graders' Proficiency in Listening and Speaking English as a Foreign Language: Focusing on Meaningful Contexts

Atsuko SHIRADO

1. Introduction

In this study, a listening test and an individual interview were used to assess students' proficiency in the English language. The tools were implemented based on students' presentations as end-products of the Project-Based Approach (PBA), which aims at evaluating students' language proficiency in meaningful contexts. A total of 70 Japanese sixth-graders participated in the study during the 2019 academic year, in the transition period to the Japanese government's new educational guidelines. The results of this study contribute to improving the evaluation methods to assess fifth- and sixth-graders' proficiency in English as a foreign language.

1.1. Background of the study

The new guidelines of the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT, 2017a) for elementary schools have established that, as of 2020, fifth- and sixth-graders must learn English as a compulsory subject called "Foreign Language." The instruction has to include 70 lessons each year and emphasize four key skills, that is, listening, reading, speaking, and writing. The main objective of Foreign Language, according to the new guidelines, is to foster students' basic knowledge and abilities to communicate in English through the four skills. The new guidelines mention three specific aims. The first aim is to acquire the four basic skills for communication in English. The second one is to encourage students' abilities to express their thoughts and feelings in English, according to the purpose, scene, and situation of communication. The third aim is to promote an independent

attitude toward communication in English, understanding the cultural differences, and empathizing with people. Therefore, the learning should not be oriented only to attain knowledge and skills of English but also to foster the ability to think, judge, and express one's ideas in English as well as developing a positive attitude toward communication and learning.

Although the new guidelines emphasize the relevance of promoting students' positive attitudes toward the English language, teachers are still required to evaluate students' progress in basic knowledge and abilities to communicate in English, examining whether the teacher's instruction and materials are appropriate for their students. Moreover, the teachers in charge of Foreign Language, who are usually homeroom teachers (HRTs), have to evaluate their students' English proficiency using a numerical three-grade evaluation on their report cards, as English has become a compulsory subject in 2020. However, the efficiency of this evaluation method of Foreign Language is debatable, as it is likely to lead to an increase in the number of students who dislike the subject. Indeed, the numerical three-grade evaluation is used as a summative assessment and tends to focus on the results of learning rather than the process itself.

Furthermore, to evaluate students' language proficiency objectively, teachers may give students some types of tests, such as written tests and listening tests, without any meaningful contexts. Although the numerical evaluation is convenient for the teachers, its use does not allow them to assess students' communication abilities adequately, as these abilities tend to emerge and be observed in meaningful contexts. As reported in the government's guidelines (MEXT, 2017a), the evaluation should address the abilities for students to express their thoughts and feelings in English in accordance with the purpose, scene, and situation of communication.

To prevent the increase in the number of students who dislike the English subject, MEXT (2017a) has recommended the use of multiple evaluation methods for Foreign Language. MEXT (2017b) has proposed not only written tests but also listening tests and performance tests, such as speeches, presentations, and interviews on authentic topics, in addition to conventional methods, such as observation and questionnaires. Teachers need to evaluate students' attitudes when they try to express their thoughts and feelings during the communication as well as students' knowledge and abilities for communication, understanding the cultural differences, and empathizing with people. This must be the method to evaluate

students' communication proficiency in meaningful contexts. Additionally, teachers should consider ways not to give an excessive load to their students in carrying out these numerical evaluation methods (MEXT, 2014). How should teachers evaluate students' proficiency in meaningful contexts? MEXT has not indicated enough practical methods yet.

1.2. The aim of this study

As an attempt at developing multiple evaluation methods, this study investigates whether a listening test and an individual interview focusing on meaningful contexts are effective to assess sixth-graders' communicative abilities.

The reasons to select these two evaluation tools are as follows. First, they allow teachers to evaluate students' listening and speaking performance objectively. Second, they enable teachers to integrate instruction and evaluation, as the latter must be carried out in the range of instruction (MEXT, 2017a). In short, teachers should evaluate students' proficiency in what teachers have already taught. Standard tests, such as the English proficiency test, cannot accomplish this aim. Third, the suggested evaluation tools allow students to think about the answers through meaningful contexts. Meaningful contexts are absolutely necessary for students to express their thoughts and feelings in English according to the purpose, scene, and situation of communication, which is the second aim of the governmental guidelines for Foreign Language. To help the students develop the abilities to express their thoughts and feelings appropriately, the teachers must foster students' listening and speaking skills through meaningful contexts during each lesson. Thus, the full integration of instruction and evaluation can be achieved using these listening tests and individual interviews.

To create meaningful contexts for each lesson, the researcher introduced the PBA into the Foreign Language curriculum with *We Can! 2* (MEXT, 2018), which is a standard material for sixth-graders created by the government and is the base of the authorized textbooks used in 2020.

The structure of this paper is as follows. Section 2 presents a literature review. Section 3 describes the study methods. Section 4 illustrates the results of the study including quantitative and qualitative data about the listening test and individual interview. Section 5 discusses the results and section 6 concludes the paper.

2. Literature review

2.1. Previous studies

The PBA is based on a theory developed by the American educational leader, John Dewey, in the early 20th century (Beckett, 2006). In general, PBA is “a collection of sequenced and integrated tasks” (Nunan, 2004, p. 133), all culminating in an end-product as the project’s core element. In the PBA process, “the route to the end-product brings opportunities for students to develop their confidence and independence and to work together in a real-world environment by collaborating on a task” (Fried-Booth, 2002, p. 6), such a process can be realized in work to improve a student’s ability to communicate in a second language (L2) / foreign language (FL). Based on the definition of the PBA in L2/FL education, empirical research on the PBA applied to English as a second language (ESL) and English as a foreign language (EFL) classrooms indicates four main features: (1) an appropriate balance between teachers’ guidance and students’ autonomy (Sheppard and Stoller, 1995; Henry, 1994); (2) the use of purposeful language (Fried-Booth, 1986, 2002); (3) multi-skill tasks (Haines, 1989; Ribé and Vidal, 1993); and (4) recycling known language (Haines, 1989). Along these lines, research advises that EFL students, who tend to have fewer opportunities to use English meaningfully, would do well in completing a collaborative project that involves discussion, research, and a presentation (Tanaka, 2009). As a result, these four aspects will allow students to increase their self-involvement, in short, an experience that nurtures feelings of self-accomplishment.

Based on the existing research reviewed above, this paper argues that English lessons based on the PBA may help teachers optimize their lessons. English lessons rooted in the PBA are designed to develop an end-product; individual vocabulary words and topics are not simply used for one lesson. Instead, they are related to one another throughout a project’s lessons. In other words, the PBA enables students to learn English most effectively by repeatedly encountering and using vocabularies and expressions related to a particular topic across different authentic contexts. This aspect of the PBA reflects the recycling of known language (Haines, 1989). Another characteristic of the PBA, the use of purposeful language (Fried-Booth, 1986, 2002), is helpful for students to develop communication abilities. If they are involved personally in an authentic context, they will try to communicate their feelings and

ideas to others with the use of purposeful language. The third feature of the PBA, multi-skill tasks (Haines, 1989; Ribé and Vidal, 1993), encourages students with different abilities to work collaboratively while playing individual important roles, as the PBA has the potential to cater to various skills within sequenced tasks of a project. This characteristic gives teachers opportunities to evaluate students from different viewpoints. That is, the PBA could generate more equal opportunities for participation and evaluation among students who have different skills and language learning styles. The last trait, an appropriate balance between teachers' guidance and students' autonomy (Sheppard and Stoller, 1995; Henry, 1994), gives students opportunities for decision-making, and requires teachers' scaffolding when students need help. This characteristic enhances students' potential and fosters their autonomy.

Shirado (2019) examined the effects of implementing the PBA in teaching the four skills of English literacy to sixth-graders of 2018 with *We Can!* 2. The study showed that the students could develop these four skills based on the questionnaires and one test for reading and writing called "Alphabet Quiz"; however, it did not use the numerical evaluation tools for listening and speaking. To complete the multiple evaluations, it is necessary to investigate how the two kinds of evaluation, the listening test and the individual interview with the assistant language teacher (ALT), affect students' evaluation. Further, the study examines whether these evaluation tools help decrease HRTs' and students' load when such evaluations are carried out, support multiple evaluations, and increase the possibility of carrying out listening tests and interviews within meaningful contexts.

2.2. Research questions

This study addressed the following three research questions (RQs).

- RQ-1: Does the listening test based on the PBA evaluate sixth-graders' listening proficiency appropriately?
- RQ-2: Does the individual interview based on the PBA evaluate sixth-graders' speaking proficiency appropriately?
- RQ-3: Do the listening test and the individual interview carried out by the ALT help decrease the HRTs' and students' load?

3. Methods

3.1. Participants

The participants were 70 students who were sixth-graders (2 classes) during the academic year of 2019 in a public elementary school in a suburb of Tokyo. They learned English through the four skills based on the PBA in school for 70 lessons during the fifth-grade, each spanning 45 minutes, and four lessons in the third- and fourth-grades, through the PBA-based listening and speaking. A preliminary survey conducted in April 2019 determined that about 21% of the participants were learning English outside of school.

3.2. Materials and design

The English lessons based on the PBA were conducted using *We Can! 2* during the first term of 2019. At the end of the first term, the listening test and the individual interview were carried out with the other evaluation tools, including two questionnaires, the alphabet quiz, and students’ and HRTs’ open-ended questionnaires (see Figure 1).

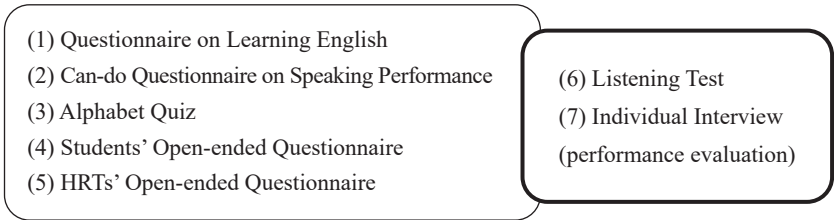


Figure 1. Evaluation Framework for Sixth-graders of 2019

Concerning the project’s design, this study employed a modified version of a project conducted in 2018, “Let’s introduce Japan and our city to the international students!” The original project was designed to explore the effectiveness of the integration of the four skills. The topic and content of the project were based on *We Can! 2*, and involved four skills. The method of the project was based on the PBA. As mentioned above, Shirado (2019) confirmed that the four-skill English lessons based on the PBA helped sixth-graders to develop the four skills during the 2018 academic year. For the current study, the design was slightly adjusted to fit better the

abilities of the students, as those in the current sample had already learned English through the four skills for 70 lessons in the fifth-grade, whereas the 2018 sixth-graders learned English just through listening and speaking in 35 lessons during the fifth-grade.

The project, "Let's introduce Japan and our city to the international students!" consisted of 22 English lessons that took place between April and July in 2019 and included vocabulary and topics from Units 1, 2, and 4 of *We Can!* 2. The goal of the project was for the students in each group to conduct research and deliver presentations with hand-made posters to six international university students in each class about the sixth-grade students' favorite Japanese events, food, and traditional games as well as places in the city recommended by the sixth-graders to the international students (see Figure 2). These international students were from the Philippines, South Korea, Taiwan, the USA, Germany, and France, and were studying at the University near the elementary school. After their presentations, the elementary school students received feedback from the international students.

This project was motivated by two goals: (1) the sixth-graders wanted to share information they collected about Japan and their city with the international students and (2) the international students wanted to learn useful information about Japan and their city from the sixth-graders.

Hello. I'm (Shirado Atsuko). My birthday is (April 11).

*(*Each student in one group introduces her / himself using these two sentences.)*

We like Japan. In (spring), we can enjoy (*hanami*).

And, we like (*dango*). It's (sweet).

Also, we like (*kendama*). It's (fun).

In our city, we have (an indoor swimming pool).

We can (swim there all year around).

*(*The students can add some more information, if they want to.)*

Please try it. Thank you for listening.

**The students can fill in the brackets with their ideas.*

Figure 2. Model Presentation

3.3. Procedure

To examine the three RQs based on the new government guidelines, this project was mainly implemented through team-teaching. Team-teaching is useful to create an authentic communicative scenario because the teachers can show the students examples of communicative interactions in meaningful contexts. Ten team-teaching sessions were carried out by the HRT and the ALT, eight team-teaching sessions by the HRT and the researcher as a Japanese teacher of English (JTE), and four solo-teaching sessions by the HRT. During the team-teaching, the HRT was responsible for the review and class management. Additionally, the HRT played the role of an exemplary English user. The JTE and ALT introduced new words and expressions and provided language support with the HRT. Further, the teachers collaboratively gave opportunities to the students to listen to the teachers' small talk about the topic of each lesson with key sentences, showed the students how to play games and exchange their feelings and ideas, and encouraged them to interact using the English language.

About one week after the presentations to the international students, the teachers (the HRT, the JTE, and the ALT) carried out the listening test and individual interviews respectively, which were used for performance evaluation. To integrate instruction and evaluation, the listening test and the individual interview were designed based on students' presentations, as shown in Figure 2. The content of the test and interview was related to what the teachers taught in the project. Moreover, the researcher, who designed these tools, noted some communicative exchanges in meaningful contexts in which the students could answer the questions by gathering information together and guessing the answers in the contexts. These two evaluation tools were implemented mainly by the ALT in different English lessons.

Given the importance of meaningful contexts for the listening test, the researcher designed it such that the students could answer the questions getting information through the flow of speech. In other words, students often could listen to more than two sentences to answer one question, gathering information from the target sentence and the sentences before and after it. When designing the test, the researcher discussed the content of the listening test, how to implement it, and the appropriateness of the English sentences with the ALT and the sixth-grade HRTs. The researcher also asked the ALT to read the script of the listening test because the students were familiar with the ALT's pronunciation and speaking style, thus

reducing the students' and the HRTs' anxiety about the test. The ALT agreed to the request and the researcher and the ALT also discussed how to read the script. For example, after the target sentence, the ALT paused the reading because the students had to draw a line.

As for the individual interview, the researcher designed a rubric for the interview and discussed the appropriateness of the rubric and how to use it with the HRTs and the ALT. In particular, all of the teachers had to agree on the interpretation of a meaningful sentence in the rubric (see Table 1).

Table 1. The Rubrics of the Interviews

Content	A student can speak a meaningful sentence without any help.	3
	A student can speak a meaningful sentence with some help.	2
	A student can say only some keywords.	1
Paralinguistic factors (clear voice and eye-contact)	Very good	3
	Good	2
	Not good	1

* Meaningful sentence: You can understand the meaning of the sentence, even if it may not be accurate grammatically. (e.g., My birthday is September seventeen.)

3.4. Data collection

This study focused on the listening test and the individual interview ((6) and (7), respectively, in Figure 1.) As the study employed a mixed methods, the quantitative data were collected using (1) the questionnaire on learning English, (2) the can-do questionnaire on speaking performance, (3) the alphabet quiz, (6) the listening test, and (7) the individual interview, whereas the qualitative data were obtained from (4) the students' open-ended questionnaire and (5) the HRTs' open-ended questionnaire. Based on the RQs, the data concerning (6) the listening test are collected and analyzed for RQ-1, and those about (7) the individual interview are done for RQ-2. The data collected for (4) the students' open-ended questionnaire and (5) the HRTs' open-ended questionnaire as well as those for (6) and (7) are useful to address RQ-3. Further, (1) and (2) are oriented to reply to RQs 1 and 2.

3.4.1. Listening test

As in the students' presentations (see Figure 2), in the listening test (see

Appendix 1) each imaginary speaker delivered a speech about her/his favorite Japanese event, food, and traditional game as well as her/his favorite place in her/his city after a self-introduction including her/his birthday. The students selected the appropriate pictures, dates, and Roman characters. They drew lines connecting the selected options. This methodology was familiar to the students because *We Can! 2* used it for the listening activities. At the end of the listening test, the students summarized each speech in Japanese by reviewing their answers on the test paper. This method was familiar to them too as also used in *We Can! 2*. The number of questions answered by drawing lines was eight for each speech and 24 in total, summarizing each speech, the questions were three in total. Each speech of the listening test was repeated twice. The students were 65, as five students were absent from the English lesson on that day. It took about 15 minutes for the students to finish the test. They needed about ten minutes to answer the questions by drawing lines, and were given only five minutes to finish summarizing each speech.

3.4.2. Individual interview

Before the interview, the JTE explained about the rubric (Table 1) and the topic of the interview. As with the listening test, the ALT carried out the individual interviews to decrease students' anxiety, as recommended by MEXT (2014). The ALT and the JTE evaluated jointly the students' interviews out of the classroom using the rubric. The HRT observed the students in the interviews, taking care of the other students who answered the questionnaires ((1) and (2) of Figure 1) in the classroom. The students were 68, as two people were absent from the English lesson. The questions of the interview were about the students' presentations: each student's birthday, favorite Japanese food, or favorite Japanese traditional game, and favorite place that she/he recommended to the international students. Before these three questions, the ALT asked each student her/his name as a warm-up. Each individual interview lasted about two-three minutes.

4. Results

4.1. Listening test

To investigate RQ-1, the data of the listening test were analyzed using percentages for achievement degree and descriptive statistics. The averages of two

kinds of achievement degree were 96% (drawing lines) and 75% (summarizing speeches) (see Table 2). As shown in Table 3, the range of standard deviation (SD) for the questions answered by drawing lines was not wide (1.50), unlike that for the questions of summarizing speeches (6.20). The mean (M) for the questions answered by drawing lines was also higher than that for the questions of summarizing speeches.

However, 48 students out of 65 (75% of the students) were able to summarize each speech with over three appropriate Japanese sentences (see Table 2). Moreover, 38 students out of 65 were able to add peripheral information which the students could not obtain just by reviewing their test papers. They seemed to understand such information about each speech only by listening, not looking at the test paper. Nevertheless, the peripheral information helped the students be convinced of the answers. For example, the students were supported to select the picture of a park, but they could not establish whether there were flowers because the picture of a park did not show any of them; thus, the ALT read, “We have a beautiful park. We enjoy watching many flowers.” In the summaries for that speech, the following sentences were found in the students’ test papers: “There are many flowers in the park,” “The flowers in the park are beautiful,” and “We can see many flowers.” Therefore, it seems that the students understood the content of each speech in the meaningful contexts.

Table 2. Average and Achievement Degree of Listening Test

N=65	Drawing Lines (24 questions)	Summarizing Speeches (3 questions)
Average	23.35/24	more than 9 sentences : 48/65
Achievement Degree	96.0%	75%

Table 3. Descriptive Statistics of Listening Test

	Frequency	Minimum	Maximum	M	SD
Drawing Lines	65	18.00	24.00	23.35	1.50
Summarizing	65	4.00	24.00	14.44	6.20

4.2. Individual interview

To investigate RQ-2, the data of the individual interviews were analyzed using percentages for achievement degree and descriptive statistics. After the interviews,

the two evaluators (the ALT and the JTE) compared their evaluation sheets each other. As a result, they found nine differences in 340 items (3%) about the evaluation based on the rubric. Each of the different evaluation items was modified by averaging between the both evaluators' figures. As reported in Table 4, the averages of the two kinds of achievement degree were 98.2 % (content) and 97.7% (paralinguistic factors).

Table 4. Average and Achievement Degree for the Individual Interviews

N=68	Content (3 questions)	Paralinguistic factors*
Average	8.84/9	5.87/6
Achievement Degree	98.2%	97.7%

*Paralinguistic factors are clear voice and eye-contact.

Table 5 shows that both ranges of SDs for content and paralinguistic factors were narrow. The means for the both were also quite high. Thus, the students could answer the questions appropriately with eye-contact and a clear voice during their interviews. At the same time, a ceiling effect may have occurred in the individual interview due to the narrow range of the SDs.

Table 5. Descriptive Statistics of Individual Interview

	Frequency	Minimum	Maximum	M	SD
Content	68	7.00	9.00	8.84	0.44
Paralinguistic factors	68	4.00	6.00	5.87	0.38

4.3. Students open-ended questionnaire

As for the listening test, some students wrote that they felt relieved because the ALT spoke to them. Given that the speeches in the test were similar to their presentations, most of the students could guess the flow of the speeches and the speech contexts. Moreover, some other students reported that because they could often listen to more than two sentences for one question, they could answer confidently.

As for the individual interview, some of the students reported that although they felt nervous, they were very happy because they succeeded in answering each question. A few students even commented that they enjoyed their interviews with the ALT.

4.4. HRTs' open-ended questionnaire

As for the listening test, one HRT wrote that the students seemed to concentrate on listening without anxiety as they listened to the ALT's speech, rather than a CD, thus understanding better the content of the ALT's speech. The other HRT commented on the listening test as follows: "Because the flow of the speeches of the test is the same as that of the students' presentations, they could guess the next topic which each speaker was speaking about and they took the test calmly." The content of each speech was different, but the topic was the same. It is effective to evaluate the students' comprehension of what the teachers taught in the English lessons. Moreover, the HRTs wrote that it was supportive for them that the ALT read the script of the listening test because the HRTs did not have to record the speeches for the test on a CD of the test. Additionally, they explained that they could entrust the ALT with the implementation of the listening test because the ALT was familiar to the students. They insisted that although the listening test implemented by the ALT was a new evaluation method, it was more effective and less mentally heavy for both the students and the HRTs.

As for the individual interview, the HRTs reported that, at first, they were worried about the interview because it was the first time for the students and would be taken individually. However, after finding that this interview was designed based on the students' presentations, the HRTs changed their minds because they knew that the students researched these topics, discussed the content of their presentations enthusiastically, and devised the ways of presentation. The HRTs thought that many of the students would answer pleasantly and confidently when they were asked about their presentations. They also thought that it was a good chance for the students to communicate with the ALT individually and impromptu. The HRTs heard some students' comments after the individual interview. For example, "It was easy," "It was interesting," or "I did it well." That was because their communication happened in meaningful contexts. Besides, based on their observations, one HRT wrote that this interview was useful for the students' learning because they could get a sense of accomplishment and because, at the same time, they could find their next goal for learning English. The HRTs thought that like the listening test, the individual interview was an effective performance evaluation tool to integrate teaching and evaluation. Furthermore, they wrote that since the interview was implemented by the ALT, the HRT did not feel any burdens.

5. Discussion

5.1. Response to RQ-1

Regarding RQ-1 (“Does the listening test based on the PBA evaluate sixth-graders’ listening proficiency appropriately?”), the results of the achievement degree and descriptive statistics of the listening test illustrate that the students’ listening proficiency improved. Although the researcher did not get the data of the students’ listening proficiency before the first term, it is considered that their improvement was appropriately evaluated. First, it should be focused on that the achievement degree of 24 questions was very high (96%). The results of the questionnaire on learning English (see Figure 1) between pre- and post-questionnaires backed up this assertion, because they also showed the increase in the students’ self-confidence and their positive attitudes toward English lessons including the four skills. The meaningful contexts in the listening test helped the students understand the content of each speech. That is why the students could gather the information before and after the target sentence to answer each question. Indeed, 75% of the students were able to summarize each speech with over three appropriate Japanese sentences. Besides, since 38 students were able to add peripheral information and summarize the speeches in detail, the meaningful contexts are supposed to have worked well.

Next, the appropriateness of the test to evaluate the students’ listening proficiency is discussed. This test was designed based on the students’ presentations. The students’ presentations were the goal of the project in the first term. In short, the presentations were condensed from the teaching in the first term, thus exemplifying the integration between teaching and evaluation. In addition, English teaching was based on the PBA. As mentioned above (see 2.1), a characteristic of the PBA, the use of purposeful language, can be accompanied by an authentic context, in which students can increase their self-involvement. In the listening test, as the test was based on the students’ presentations, they were likely to empathize with each speaker of the test and be involved in the speech personally. That is, the context of the test was authentic and helpful for the students. Thus, this listening test could evaluate the students’ listening proficiency in an appropriate situation.

Moreover, the scores of the listening test, which are a form of numerical evaluation, can be used as an indicator to represent the students’ present listening abilities objectively. Therefore, the answer to RQ-1 is positive.

5.2. Response to RQ-2

Concerning RQ-2 (“Does the individual interview based on the PBA evaluate sixth-graders’ speaking proficiency appropriately?”), the results of the achievement degree and descriptive statistics of the individual interview demonstrate that the students’ speaking proficiency developed. Although the researcher did not get the pre-data of students’ speaking proficiency before the first term, it is thought that their development of speaking proficiency was evaluated appropriately. As for the development of the students’ speaking proficiency, first, the achievement degrees were very high (see Table 4). Besides, the results of the can-do questionnaire on speaking performance (see Figure 1) between pre- and post-questionnaires supported the data of the individual interviews, because they also showed the increase in students’ self-confidence about speaking English. That is because the individual interview created an authentic context with the use of purposeful language, a characteristic of the PBA. The authentic context encouraged the students to communicate in English personally and impromptu. Moreover, like the listening test, the topic of the individual interviews was about the students’ presentations, which reflects the integration of teaching and evaluation. Thus, this context could bring out the students’ speaking proficiency appropriately. The scores of the individual interview, which are a form of numerical evaluation, can be also used as an indicator to represent the students’ present speaking abilities objectively. Consequently, the answer to RQ-2 is positive.

5.3. Response to RQ-3

Concerning RQ-3 (“Do the listening test and individual interview carried out by the ALT help decrease the HRTs’ and students’ load?”), the results of the students’ and HRTs’ open-ended questionnaires showed that these two evaluation tools were helpful to decrease HRTs’ and students’ mental burdens, such as the anxiety for the evaluation. As for the listening test, some of the students wrote that because the ALT spoke to them, they felt relieved. Moreover, the HRTs wrote that it was useful for them that the ALT read the script of the listening test, because the HRT did not have to prepare a CD for the test. Additionally, they explained that they could entrust the ALT with the implementation of the listening test because the ALT was familiar to the students. They thought that although the listening test implemented by the ALT was a new evaluation way, it was very effective for the students and the HRTs and

did not represent a mental burden for either of them.

As for the individual interviews, some of the students reported that although they felt nervous, they were very happy because they succeeded in answering each question. Moreover, the HRTs heard some students' commenting that the interview was easy and interesting. The HRTs also wrote that since the interview was implemented by the ALT, the HRT did not feel any burdens. Hence, the answer to RQ-3 is positive.

This study has some limitations. First, because of educational and ethical reasons, this study did not involve a control group. Comparing experimental groups with a control group would have been useful to offer more insights into the characteristics and effects of the evaluation tools based on the PBA. Further, as a ceiling effect may have occurred in the individual interviews, it is necessary to improve this tool in future work.

6. Conclusions

This study explored how two new evaluation tools, a listening test and an individual interview, measured the English communicative abilities of 70 sixth-graders. The results of the listening test and the individual interview were discussed and yielded positive responses to the study's three research questions, suggesting that it is appropriate to adopt these two evaluation tools in elementary school for the evaluation of students' proficiency in the English language. These evaluation tools relieve HRTs from the mental burdens of the evaluation, as they are conducted by the ALT. Additionally, if these listening and speaking evaluation tools are added to conventional methods, such as questionnaires and written tests, the evaluation methods will be more suitable to examine "students' basic knowledge and abilities to communicate in English through the four skills," as recommended in the MEXT's new guidelines (2017a, 2017b). Therefore, these listening and speaking evaluation tools can support multiple evaluations, integrate teaching and evaluation, and increase the possibility of carrying out listening tests and individual interviews with meaningful contexts. Further research on the effectiveness of the evaluation methods for elementary school English education will be discussed in a subsequent study.

References

- Beckett, G. H. (2006). *Project-based second and foreign language education: Theory, research, and practice*. In G. H. Beckett and P. C. Miller (Eds.), *Project-based second and foreign language education: Past, present, and future* (pp. 3-16). Greenwich, Connecticut: Information Age Publishing.
- Fried-Booth, D. L. (1986). *Project work*. Oxford: Oxford University Press.
- Fried-Booth, D. L. (2002). *Project work* (2nd ed.). Oxford: Oxford University Press.
- Haines, S. (1989). *Projects for the EFL classroom: Resource material for teachers*. Edinburgh: Thomas Nelson and Sons.
- Henry, J. (1994). *Teaching through projects*. London: Kogan Page.
- Ministry of Education, Culture, Sport, Science, and Technology (MEXT). (2014). Kongono eigokyoikunokaizen, jyujihsuhosaku nitsuite hokoku ~ gurobaruka nitaishita eigokyoikukaikaku no itsutsunoeian [The report about the improvement and substantiality of future English education: Five proposals for English education improvement corresponding to globalization]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chousa/shotou/102/houkoku/attach/1352464.htm
- Ministry of Education, Culture, Sport, Science, and Technology (MEXT). (2017a). Shogakko gakushushidoyoryo [The plan of elementary school government guidelines]. Retrieved from http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afiedfile/2017/05/12/1384661_4_2.pdf
- Ministry of Education, Culture, Sport, Science, and Technology (MEXT). (2017b). Shogakko gaikokugokatsudo gaikokugo kenshu gaidobukku [The guidebook of study training of foreign language activities and foreign language of elementary schools]. Retrieved from https://www.mext.go.jp/a_menu/kokusai/gaikokugo/_icsFiles/afiedfile/2017/07/07/1387503_1.pdf
- Ministry of Education, Culture, Sport, Science, and Technology (MEXT). (2018). *We Can! 2*. Tokyo: Tokyoshoseki.
- Nunan, D. (2004). *Task-based language teaching*. Cambridge, UK: Cambridge University Press.
- Ribé, R., & Vidal, N. (1993). *Project work: Step by step*. Oxford: Heinemann International.
- Shirado, A. (2019). A possible project-based approach to elementary school English education: Teaching four skills to sixth graders. *Tsuda Inquiry*, No. 40, 1-19.
- Sheppard, K., & Stoller, F. (1995). Guidelines for the integration of student projects into ESP classrooms. *English Teaching Forum*, 33 (2), 10-15.
- Tanaka, S. (2009). Shogakko no eigokyoiku wo kangaeru—purojekuto gata toshiteno eigokatsudou [Considering elementary school English education—foreign language activities as projects] *The English Teachers' Magazine*, June, 60-63.

Appendix 1

聞くこと・クイズ

6 年 () 組 () 番

3 人がそれぞれクイズ形式で自己紹介を通して 2 回言います。よく聞いて内容に当てはまる絵の ● を線で結びましょう。聞いた後、3 人について分かったことを口の中に日本語で書きましょう。



Akane



Kenta



Ayumu



2月26日



5月16日



3月16日



5月26日



3月6日



えのく





Midori City

Minami City

Kodaira City

Sakura City



* Akane について分かった事

* Kenta について分かった事

* Ayumu について分かった事