The Potential of Student-Generated Questions and Answers as Scaffolding Process for Japanese EFL Students’ Reading

YONEZAKI Michi *

1. Introduction

Q&A activities are one of the traditional teaching activities used to measure students’ comprehension of reading texts. According to a survey on English instruction in Japanese secondary schools by Benesse Corporation (2015)\(^1\), 87.1 percent of junior high school teachers and 89.4 percent of high school teachers answered that they use Q&A activities ‘often’ or ‘sometimes.’ Most Q&A activities in the current classroom environment in Japan, however, are conducted with teacher-centered instruction. In typical Q&A activities, the teacher asks questions and students answer them. Therefore, rarely do students ask questions, and the number of students’ questions is limited (Graesser & Person, 1994\(^2\)). This is one reason why many students are not very good at forming questions (Nuttall, 1996\(^3\)). Furthermore, several researchers (e.g., Applegate, Quinn, & Applegate, 2002\(^4\); Farahian & Rezaee, 2012\(^5\); Long & Sato, 1983\(^6\)) pointed out that most of the types of questions teachers ask are those calling for factual details in the text, whose answers are already written in the text.

This study suggests that teachers should provide opportunities for students to try asking questions and answering them by themselves or with peers. This is because the act of asking questions and searching for answers leads to active learning, and generating their own questions is also a creative act (Chin, 2002\(^7\)). Therefore, we assume that Q&A activities might be more interesting, more challenging and better at fostering students’ cognitive and affective engagement, if we change the current style of Q&A activities (teacher-to-student Q&A) into Q&A activities conducted among students (student-to-student Q&A).

To conduct a student-to-student Q&A, this study will highlight instruction through student-generated questions. Student-generated questions encourage students to make up their own questions before, during, and/or after reading a passage. A variety of student-generated instructions and procedures have been used in various educational settings. The approaches and procedures are different for research designs and educational purposes. A lot of researchers (e.g., Janssen, 2002\(^8\); Janssen, Braaksma, & Couzijn, 2009\(^9\); National Reading Panel, 2006\(^10\); Risko & Feldman, 1986\(^11\),

---

\(^*\) 甲南女子大学（Konan Women’s University）
Rosenshine, Meister & Chapman, 1996\(^{(12)}\); Rouse, 2014\(^{(13)}\) insist that instruction through student-generated questions is an effective approach to improve students’ reading comprehension and their attitude towards learning. However, most of the previous studies (e.g., Janssen, 2002; Janssen et al, 2009; Risko & Feldman, 1986; Rouse, 2014) were conducted in L1 classroom settings, and only limited studies on EFL were conducted (to be discussed in detail later). Few studies have been reported in a Japanese EFL context as far as the current researcher knows. Thus, this study attempts to examine the effects of a student-generated questions approach to Japanese EFL students.

2. Literature Review

2.1 The Value of Fostering Student-Generated Questions

There are several reasons why this study focuses on the student-generated question approach. First, self-generated questions are assumed to play an important role in fostering students’ comprehension. When students compose their own questions, they focus their attention on the content (Rosenshine et al., 1996). Students may become more involved in reading when they generate questions and answer their own questions, rather than when they merely respond to questions from a teacher or in a text (Janssen, 2002). Furthermore, generating questions may require students to play an active and initiating role in the learning process (Rosenshine et al., 1996). When students actively engage in learning, they build more knowledge and improve their comprehension, rather than when they passively receive information (Graesser & Olde, 2003\(^{(14)}\)). This may lead to an enhancement in the students’ motivation. Furthermore, if students share their own questions and seek the answers with their peers, students’ engagement may be promoted more, which in turn may lead to a higher level of responses and deeper understanding (Nguyen, Janssen, Rijlaarsdam, Rijlaarsdam, & Admiraal, 2016\(^{(15)}\)). In other words, students may learn to ask and answer questions that require them to integrate parts of texts or make inferences, using their prior knowledge and experiences.

Second, asking questions is a natural response to reading (Janssen, 2002; Nguyen et al., 2016). When reading, texts do not always give us all the meanings. Readers may ask questions about the text and express uncertainties, to make sense of stories. This is a natural process of reading. However, rarely have students been given opportunities to ask their own questions and discuss uncertainties in the classroom, and most of the questions are asked by teachers or in textbooks. Furthermore, most of the questions are text-based questions, which call for only a shallow understanding of the text, extracting meaning from the text and recalling the information in the text. Graesser and Olde (2003) reported from Dillon’s (1988\(^{(16)}\) survey that only 4% of the questions asked by the teacher were deep questions.

Finally, the student-generated question approach encourages students’ metacognitive skills (Jin & Shin, 2012\(^{(17)}\)). This is because when students generate questions, they can identify what they have understood and what they have not. Therefore, student-generated questions during and/or after reading are “a strategy that enables students to monitor their reading comprehension and increase their ability to learn independently” (Joseph, Alber-Morgan, Cullen & Rouse, 2016\(^{(18)}\), p.152). When they become aware of what is wrong and undesirable about the questions and answers they generated, they will do something to make up for it (for example, read again and again or ask their teacher).

2.2 Relevant Research

The first training studies on student-generated questions date from the 1960s (Janssen, 2002). Since then, a lot of studies have examined the effects of instruction through student-generated questions on students’ reading. Wong (1985\(^{(19)}\)) examined a total of 27 studies published from 1965 to 1982. In a subsequent review, Rosenshine et al. (1996) conducted a review of 26 student-generated question intervention studies, published between 1983 and 1992. Janssen (2002) also reviewed 32 different self-questioning studies, published between 1992 and 2000. A more recent review was conducted by Joseph et al. (2016), published from 1990 to 2012. Each review suggested several factors contributing to the success of student-generated questions. For example, Wong (1985) concluded that successful instruction provided students with effective training including modeling, imitation, and reinforcement.
Rosenshine et al. (1996) investigated the use of successful procedural prompts. Procedural prompts are visual or auditory cuing or signaling, and can be scaffolds that are specific to the cognitive strategy. They found that teaching students to ask themselves to start with signal words (who, what, where, and how), generic question stems (e.g. How are...? / What is the main idea of ...?), and story grammar categories (e.g. setting, main character, character’s goal, obstacles) were the most successful prompts. Janssen (2002) argued that students should be encouraged to generate questions about anything that they find puzzling or that captures their interest in the text, which would generate authentic questions.

While early research on self-generated questions focused on question generation as a reading-comprehension strategy, current research includes various other strategies and procedures, such as peer-teaching. Although various types of student-generated questions were effective, Joseph et al. (2016) suggested that the key point to make instruction successful would be to "consider incorporating effective instructional components of modeling, prompting, using graphic organizers, and providing sufficient opportunities to practice with corrective feedback" (p. 171).

In sum, student-generated questions in previous research appear to be an effective approach to improving students’ text comprehension. However, most of these previous studies were conducted in L1 educational settings. In EFL educational settings, only a few studies have been conducted. Nguyen et al. (2016) examined the effects of instruction through student-generated questions on Vietnamese college students’ literary reading engagement. Two treatment groups and one control group were prepared: initially two groups received student-generated question instruction with a group discussion and with individual free writing, and one group received conventional instruction with teacher-generated questions. During the experiment the roles were switched, where the treatment group and control group were switched after the initial experiment was done. They found that the treatment groups showed positive effects. Furthermore, there were no significant differences between the two treatment groups.

Baleghizadeh (2013) examined the effect of English reading comprehension for Iranian college students, using peer-interaction through student-generated questions, in which students generated their own questions based on the passage, and then worked in pairs and took turns posing their questions to each other and answering and discussing each other’s questions. Two experimental groups and one control group were prepared. Both of the treatment groups had opportunities to discuss their own questions based on the passage with a peer. The first treatment group read unmodified text while the second treatment group read a simplified version of the same text. On the other hand, the control group read the unmodified text without any opportunity for interaction. The results showed that reading comprehension significantly improved when students had opportunities for peer-interaction through student-generated questions. Furthermore, students who read unmodified texts supported by peer-interaction outperformed those who read the simplified text.

3. Research Questions

Much previous literature concluded that training students to generate questions, during or after reading texts, is a useful strategy to interpret texts. Despite the advantages of self-questioning in reading, few studies have focused on Japanese EFL learners as far as the current researcher knows. Thus, this study focuses on Japanese EFL learners as participants.

It is said that the ability to generate questions is an advanced reading comprehension skill (Rouse, 2014). Generating and answering higher-order thinking questions, such as inferential questions or implicit questions, in particular, require students to read the text at a deeper level. Therefore, we are wondering whether our participants, including those who are not good at English, can generate higher-order thinking questions such as inferential or implicit questions by themselves. In addition, little research has been done on the quality of questions that are constructed and asked by students. In other words, most previous research did not reveal the question types that students generate. Thus, this study examines what kind of question types the participants generate.

Finally, most outcome measures in previous
research consisted of multiple choice or short answer literal and inferential comprehension questions, asked by the experimenter or standardized tests, to examine the participants’ reading skill (Joseph et al., 2016). No studies seem to analyze students’ individual responses to questions they generated themselves in class. Therefore, this study closely monitors how the participants’ responses to their own questions improve, in the light of quality and quantity. If the participants generate such responses that extend beyond the text and reflect analysis of the text, or comprehend the implications of the text, it means they understand the text deeply.

Keeping these previous research limitations and our quests in mind, our main research questions are framed as follows:

1) Can the participants, including those who are not good at English, ask higher-order thinking questions, if given some instruction?
2) Do the participants’ responses to the questions they generate improve, in the light of quality and quantity?

4. Method
4.1 Participants
A total of 36 university students learning English as a foreign language participated in this study. The participants were divided into two groups. Twelve participants, categorized as the upper group, were sophomores from the department of English Language and Culture. Their TOEIC scores ranged from 500 to 675. On the other hand, 24 participants, categorized as the lower group, were freshmen from two different departments; Child Development and Education, and Human Life Environment. They did not obtain TOEIC scores, but these participants were enrolled in a basic English level class. Each group was given English instruction separately by the current researcher once a week (90 minutes). Both groups had received no prior training on how to form questions related to texts.

4.2 Procedures
The experiment was conducted in the fall semester of academic year 2016 at a private university, where the current researcher works. About forty minutes were assigned for the experimental treatment in both the groups. The rest of the lesson hours were spent mainly on listening and speaking practice.

Both groups received 12 sessions, as shown in Table 1: 3 for training and 9 for practice. The researcher believed that just telling students to generate questions is difficult for the participants and that they should be given some instruction. Therefore, it was decided to have some training sessions. This was because the participants had no idea about generating their own questions in English related to the text. However, it is not the purpose of this study to determine what training should be and how training and practice should be implemented in the classroom.

During the first training session (T1), the instructor provided an overview of activities for using self-questioning while reading English stories and modeling the generation of questions. The question types the instructor provided were all factual. Then, the participants were encouraged to generate questions (only wh-questions, not yes/no questions) and answers in English about the story the instructor had prepared. At the end of T1, some samples of questions and answers the participants generated were presented in the class, followed, if necessary, by corrective feedback which was mainly related to grammar correction.

After T1, three sessions as Practice 1 (P1) were followed. The instructor assigned short stories and asked the participants to generate three questions and answers before class each week. The participants brought the questions they formulated to the class. After a brief explanation of the background information of the story by the instructor and various types of oral reading in class, the participants took turns posing their questions to each other and answered them in pairs. At the end, they submitted a sheet comprising the three questions and their answers. The sheet was returned in the next lesson with the instructor’s corrective feedback or comments, if necessary.

In the second training session (T2), the instructor gave a list of question types (factual questions, low-level inferential questions, high-level inferential questions; see 4.4.1) and examples of each question type. The participants were encouraged to answer each type of question the instructor posed.
Then they discussed which type of questions was more challenging and interesting for them. They also discussed why some question types were more challenging and interesting than others. Then they were told to categorize the questions they had generated before, into three question types.

After T2, three sessions as Practice 2 (P2) were followed. Students were not mandated to make each of the question types demonstrated in T2, but were encouraged to generate questions and answers they wanted to ask. In P2, student-generated question and answer activities were conducted in the same way of P1.

Peer-questioning discussion was introduced in the last training session (T3). In peer-questioning discussion, the participants worked in small cooperative groups and took turns posing their questions to each other and answering each other’s questions. Then they were encouraged to evaluate and compare their answers, and if necessary, to correct them. After T3, three sessions as Practice 3 (P3) were conducted. In P3, the peer-questioning discussion was added; they chose two or three questions for discussion which they felt were most challenging, interesting or inspiring, and shared their answers or found better answers in small groups. The participants submitted a sheet and the instructor gave corrective feedback and some comments, if necessary.

4.4.1 Question types

To capture transition in the question types among practice sessions, we analyzed the question types the participants generated on self-questioning sheets in the light of a framework from Applegate et al. (2002): factual questions, low-level inferential questions, high-level inferential questions, and response questions. Factual questions are answered directly and explicitly from the text. They simply require that learners recall what they have read. An example of factual questions is as follows:

Factual question:
The text states that Mary, a character in the story, is in the sixth grade. A factual question asks, “What grade was Mary in?”

Inferential questions require learners to put together pieces of information that are scattered throughout the text. Learners may draw a conclusion based on the text and sometimes combine their literal understanding of the text with their own knowledge.
and background experience. The answers to low-level inferential questions are not stated directly in the text but are relatively obvious. On the other hand, high-level inferential questions require learners to use significantly more complex thinking than low-level inferential questions, because learners may need to link their own experience with the text and draw a logical conclusion. To distinguish between low-level and high level inferential questions, here are some examples from Applegate et al. (2002):

Low-level inferential question:
The text states that Mr. Wilson’s car would not start and Mr. Wilson was late for work.
A low-level inferential question asks, “Why was Mr. Wilson late for work?”

High-level inferential question:
The text describes two characters and several circumstances in their lives.
A high-level inferential question asks, “Why do you think that the two characters in the story became friends?”

Response questions require learners to express their feelings, ideas, or evaluations about the text. The answers are not found in the text and they come from the learners, although they must relate to the content of the text. This study initially did not intend to include the response question type before the experiment, because response questions can be asked without reflecting on the text and may be easier when readers ask a question such as “how did you like this story?” Therefore, we did not show participants examples in T2. However, some participants generated response questions. Therefore, we decided to include this type in our analysis.

Factual questions are categorized as lower-order thinking questions, while inferential questions have been termed as higher-order thinking questions (Lesilie & Caldwell, 2009). While low-level inferential questions demand some higher-order thinking, they have relatively obvious answers, as we mentioned above. Therefore, this study does not regard low-level inferential questions as higher-order thinking questions. In addition, as we mentioned above, some response questions can be asked without

---

Table 2. Response scales (from Bradley et al., 2008)

<table>
<thead>
<tr>
<th>Score</th>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No score</td>
<td>Students attempted submission but their answers were incorrect or too off-topic.</td>
</tr>
<tr>
<td>1</td>
<td>Reading citation</td>
<td>Students only cited the reading, using mostly direct quotations to answer their question.</td>
</tr>
<tr>
<td>2</td>
<td>Content clarification</td>
<td>Students reorganized or paraphrased ideas in their own words.</td>
</tr>
<tr>
<td>3</td>
<td>Personal interpretation</td>
<td>Students’ answers were rather a shallow understanding of the story although they applied a personal experience or scenario to answer their question.</td>
</tr>
<tr>
<td>4</td>
<td>Making inferences</td>
<td>Students’ answers reflected analysis, synthesis or evaluation, or deep understanding of the implications of the story.</td>
</tr>
</tbody>
</table>

---

Table 3. Transition of question types

<table>
<thead>
<tr>
<th>Practice Session</th>
<th>Factual</th>
<th>Low-level inference</th>
<th>High-level inference</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper</td>
<td>Lower</td>
<td>Upper</td>
<td>Lower</td>
</tr>
<tr>
<td>P1</td>
<td>87</td>
<td>183</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>87.0%</td>
<td>84.7%</td>
<td>11.0%</td>
<td>13.9%</td>
</tr>
<tr>
<td></td>
<td>2.0%</td>
<td>1.4%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>P2</td>
<td>58</td>
<td>137</td>
<td>28</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>54.2%</td>
<td>66.2%</td>
<td>26.2%</td>
<td>23.7%</td>
</tr>
<tr>
<td></td>
<td>15.9%</td>
<td>9.7%</td>
<td>3.7%</td>
<td>0.5%</td>
</tr>
<tr>
<td>P3</td>
<td>15</td>
<td>92</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>15.3%</td>
<td>44.4%</td>
<td>32.7%</td>
<td>18.4%</td>
</tr>
<tr>
<td></td>
<td>45.9%</td>
<td>36.7%</td>
<td>6.1%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>
adequate understanding of the text. Therefore, although some researchers categorize response questions as higher-order thinking questions (e.g., Applegate et al., 2002), this study regards only high-level inferential questions as higher-order thinking questions.

A total of 935 answers were categorized into question types by two researchers. When it was difficult to judge which category a question belonged to, the current researcher and the other cooperative researcher discussed this and reached agreement.

4.4.2 Analysis of answers

To examine the quantity and quality of improvement of the participants’ answers, we analyzed the participants’ answers to the questions they generated on self-questioning sheets in two practice sessions: Pre (the first session of P1), and Post (the last session of P3). Word count was employed to examine the improvement of the quantity. This is because word count is easy to assess, is an objective measurement of student responses (Bradley, Thom, Hayes & Hay, 2008), and is used as a quantitative measure of students’ participation (Andrews, 1980). As for the quality of improvement, we adapted the scheme of Bradley et al. (2008) (Table 2), in which students’ answers were categorized according to the thinking level, based on the taxonomy of Bloom (1956). The lowest score of zero was incorrect answers in Bradley et al.’s (2008) scheme. The highest score of four, which was considered as higher-order thinking, related to students making inferences by analyzing, synthesizing, or evaluating the text.

Each answer received only one code, and if there were two codes included in the answer, the highest code was earned. Two researchers analyzed the participants’ answers. Inter-rater reliability between them was $r = 0.92$. Therefore, the evaluation by each evaluator was adopted as it was.

5. Result

5.1 Question Types

The numbers and percentage of question types that students generated during experimental treatment were calculated (see Table 3). Figure 1 graphically presents the result of the ratio of upper group. Figure 2 presents lower group.

At the beginning of the experiment, the majority in both groups generated factual questions. In addition, there were many errors of question forms, even in the upper group. However, as more practice sessions were conducted, more low- or high-level inferential questions were generated, instead of factual questions, in both groups. At the end of the experiment, the high-level inferential questions in the upper group outnumbered the factual questions. Although the participants in the lower group still asked many factual questions at the end of the experiment, the number of high-level inferential questions also increased, which is noteworthy. Although we did not show the model of response questions in P2, some participants in the upper group generated response questions.

To examine the effects of treatment in detail, chi-square tests were conducted in each group. The frequency of four question types in each Practice session (P1, P2, P3) was computed. As the result of chi-square tests, significant differences were found in both groups: the upper group ($\chi^2 (6) = 111.25, p = 0.00, \text{Cramer’s V} = .43$), the lower group ($\chi^2 (6) = 123.88, p = 0.00, \text{Cramer’s V} = .31$). In addition, significant standardized residual analysis was
performed. Results are shown in Table 4 (upper group) and Table 5 (lower group).

As for the upper group, the frequency of factual questions in P1 was significantly larger, while the frequency of other question types was significantly less. However, in P3, the frequency of factual questions was significantly reduced and low- and high-level inferential questions were significantly increased. As for the lower group, the result seemed to be similar to the upper group’s, but it took more time for the lower group to be able to ask high-level inferential questions, compared to the upper group. In P2, the frequency of low-level inferential questions significantly increased, but the frequency of high-level inferential questions was still less. However, in P3, the frequency of high-level inferential questions significantly increased and the factual questions significantly decreased.

These results show that as more training and practice sessions were conducted, although the time taken depended on the learners’ level, both groups increased the number of high-level inferential questions.

### 5.2 The Quality and Quantity of Answers

To examine the quantity of the participants’ answers, we computed the average word count on the participants’ answers to each question. Table 6 shows the descriptive statistics of Pre and Post sessions. The average word count increased in both groups and significant differences were also found between the Pre and Post in both groups (upper group: \( t(10) = -8.31, p = 0.00, d = 3.50, 95\% \text{ CI} [-13.14, -7.59] \), lower group: \( t(22) = -7.77, p = 0.00, d = 2.34, 95\% \text{ CI} [-1.61, -0.94] \)). The effective size for this analysis of both groups \((d = 3.50, d = 1.68 \text{ respectively})\) was found to be large \((d = .80)\). This result indicates that both groups generated longer responses and their quantity improved after the treatment.

To examine the improvement of the quality of the participants’ answers, the answers to each question were categorized into five different codes (higher-order thinking level) based on Bradley et al.’s (2008) scheme (Table 2), and the average score of each answer was calculated. The result (Table 7) shows that the average scores increased in both groups after the treatment, and significant differences were also found in both groups (upper group: \( t(10) = -12.82, p = 0.00, d = 4.26, 95\% \text{ CI} [-2.19, -1.54] \), lower group: \( t(22) = -7.77, p = 0.00, d = 2.34, 95\% \text{ CI} [-1.61, -0.94] \)). The effective size for this analysis of both groups \((d = 4.26, d = 2.34 \text{ respectively})\) was found to be large \((d = .80)\). This result indicates that both groups generated more complex and varied answers and the quality of answers also improved after the treatment.

### To show how their answers improved, here are some samples of participants’ questions and answers in each group generated in P1 and P3. The samples are quoted as they were:

**Participant A (in lower group)**

[P1]

Q: Where was the author born?

A: She was born in Tokushima.

[P3]

Q: Why did Ben Price leave the bank?

A: Because Ben Price forgave Jimmy. He probably did not want to break Jimmy’s happiness.
In P1, the question was a factual one and the participant generated the answer just by extracting it from the text, as it was. However, in P3, the participant asked a high-level inferential question and answered it using personal interpretation.

Participant B (in upper group)

[P1]
Q: What was the author good at?
A: She was good at playing the piano.

[P3]
Q: Why did Jimmy smile sadly?
A: He knew his life with Annabel was over, when he broke into the safe. Annabel would know that he was a safe breaker. He wanted to tell Annabel his feelings, but he couldn’t. He can’t help smiling sadly.

Participant B’s question in P1 was also a factual question and the answer was extracted from the text. In P3, the participant asked a high-level inferential question and the answer reflected making the inference and became longer and more complex, although some grammatical errors were found.

Some participants generated question types categorized in the high-order thinking level in P3, but some of their answers did not lead to a higher level of response because of a lack of inference, misinterpretation of the story, or limitations of English. Examples are furnished below:

[P3]
Q1: Why did Jimmy smile sadly?
A1: Because he made a promise to himself but he broke it.

Q2: How did Jimmy feel when he didn’t look back?
A2: He thought he won’t meet Annabel again.

As for A1, the participant just extracted a sentence from the text and the answer lacked inference. On the other hand, A2 was an example of a shallow understanding of the text and a deeper understanding and more implication of the story were necessary.

6. Discussion

The first research question was “Can the participants, including those who are not good at English, ask higher-order thinking questions, if given some instruction?” Most of the questions the participants in both groups generated belonged to the factual category at the beginning of the experiment. This was probably because the participants were not used to generating questions and it was easier for them to pick factual ones. In addition, most of the question types the teachers or the textbook asked in the classroom were factual questions. The modeling of the generation of questions in T1 may have also influenced the participants. However, after the three training sessions and several practice sessions, higher-order thinking questions gradually increased. In the upper group, more high-level inferential questions were generated compared to factual questions. To generate high-level inferential questions, students need to go beyond the text as well as draw more inferences, and use their background knowledge to make sense of the text (DuBravac & Dalle, 2002). This process seemed to be difficult for some participants in the lower group. In fact, more factual questions were generated than high-level inferential ones in the lower group, even in P3. However, it should be noted that, after the treatment, even the lower group generated high-level inferential questions.

One of the possible reasons why they generated higher-order thinking questions is that the participants may have learned how to generate higher-order thinking questions through the training and practice sessions. Although the purpose of the study is not to determine what instruction should be done, these trainings and practices may have contributed to these results. Further study needs to determine what trainings and practices should be and how they should be implemented in the classroom.

Another possible reason why the participants generated higher-order thinking questions is that the participants may have realized that higher-order thinking questions were more challenging and difficult, but worth discussing. Although students were not mandated to make each of the question types, more higher-order thinking questions were generated by them. While training and practice continued, they may have found lower-order
thinking questions boring, since these questions only require the participants to restate the facts in the text. On the other hand, to generate and answer higher-order thinking questions, they had to make additional efforts to read the text, find key ideas and relationships, and integrate them into their prior knowledge. The participants may have found this process more challenging, no doubt, but at the same time, interesting or worth discussing. In fact, most of the participants asked the higher-order thinking questions when peer-questioning discussion was conducted. However, whether the participants thought the higher-order thinking questions were interesting or worth discussing was not verified in this study. Therefore, further investigation is needed.

The second research question examines whether the participants’ responses lead to a higher level of responses through student-generated question activities. The results showed that participants’ responses improved both in quality and quantity. Two reasons for these results were considered. First, the question types generated by participants reflected their answers. The number of high-level inferential questions increased as the sessions went on. Answers to high-level inferential questions include more complex, creative and varied answers. Many of the answers to high-level inferential questions were coded as “making inferences” on Table 2, although some of them were judged as “personal interpretation” or “no score” when the answers lacked inference or were incorrect. On the other hand, answers to lower-order thinking questions could simply be extracted from the textbook. Most of the responses to lower-order thinking questions were coded as “reading citation” on Table 2. Therefore, the length and variety of participants’ responses were limited.

Another possible reason is that pair work and peer-questioning discussion may have contributed to this result. During the pair work and peer-questioning discussion, some participants shared each other’s ideas and opinions, cooperated to work on finding answers, and received support and encouragement from peers. On the other hand, other participants experienced some peer pressure, which forced them to propose good questions and to provide elaborate answers. Janssen (2002) pointed out that peers’ models represent more effective instruction than teachers’ models, which may apply to our research, too.

7. Conclusion

Although our research has just started to examine whether instruction through student-generated questions can be used effectively for Japanese university students, the findings have several implications for practice. Such instruction enabled our participants, including those who are not good at English, to ask more questions based on higher-order thinking and to learn to generate higher level responses, after the treatment. This means that students may engage in reading and comprehend the text better. Since we decided that telling the students to generate questions was not enough and that they needed time and practice, we showed and explained model questions to the students and gave opportunities for peer questioning and discussion.

In Japan, the government curriculum guideline has emphasized an active-learning approach and independent learning. Students’ engagement and personal development of learning as a language learner have become the focal point of foreign language education. Instruction through student-generated questions will have the potential of enhancing students’ engagement in learning and of promoting independent learning. If instruction through student-generated questions is a useful strategy, the teachers’ role may shift from providing literary knowledge to coaching learners’ individual reading processes (Janssen, 2002).

There are some limitations to this study. It was conducted with a small number of participants. Because of this, we analyzed the participants’ questions and responses. However, we need to conduct further research with a larger sized sample and in various educational settings.

Secondly, as for the reading material used, this study prepared three short stories considered suitable to our participants. Although this study used similar types of texts including length and reading difficulty, the results may have been dependent on this text type. Therefore, further research should be conducted with other reading material, for example, expository texts, or more advanced texts with various dimensions for interpretation.

Thirdly, regarding the experiment design, the study considered the first session of P1 as Pre.
However, the Pre occurred after the initial training in questioning technique because we thought that the participants had no idea about generating their own questions in English related to the text. The result might have changed if we had conducted the Pre session before training.

Finally, regarding the measurement of participants’ outcome, this study analyzed the participants’ responses to the questions they generated. Research with another measure should be conducted, for example, assessing responses to reading comprehension questions in a different text, standardized reading assessments, retelling the main idea of the text, having students write a summary of the text, etc.

Although several limitations exist, we hope that our findings will have significant implications for teaching Japanese EFL students.

Note
1. Three textbooks were used: VISTA English Communication I (Sanseido), Vivid English Communication II (Daiichi Gakushusha), VISTA English Communication II (Sanseido)

References


