The Effects of Questions on EFL Learners’ Situation Models: Types of Question, Text Levels and Learners’ L2 Reading Proficiency

Masahiro MAEDA
Kanazawa Commercial High School

Abstract
The present study investigated the effects of three types of questions—fact-finding questions (FFQ), backward inference questions (BIQ), and self-reference questions (SRQ)—on learners’ construction of situation models during reading instruction in an English-as-a-foreign-language (EFL) environment. In this experiment, 89 Japanese high-school students read two passages with different text levels, and their comprehension and construction of situation models were assessed via a free recall task (FRT) and a sentence recognition task (SRT). For FRT, the recall data was analyzed qualitatively as well as quantitatively. The results indicated that the participants in the upper group did not generate inferences without questions in the difficult text, but they generated them without questions in the easy text. However, participants given SRQ in the upper group generated inferences even in the difficult text. On the other hand, FFQ and BIQ did not activate the generation of inferences by the participants in the upper group in the difficult text. In addition, we found that there is a possibility that questions would sometimes interfere with the generation of inferences as well as surface-level comprehension in the easy text in the upper group. The results also indicated that the generation of inferences was not activated for the participants in the lower group without questions, but inferences were generated if they were given SRQ.

Keywords: situation models, reading instruction, reading questions, text levels, L2 reading proficiency

Questions in the Instruction of Reading Comprehension
How many and what kinds of questions should teachers ask to help learners comprehend texts well and deeply? What does it mean for learners to comprehend texts? Quite a few teachers have tried to find the answers to these questions in teaching reading, but it has never been fully clear how teachers can foster text comprehension through questions and how teachers can evaluate when learners sufficiently comprehend texts. In order to comprehend texts, learners need to use top-down and bottom-up reading skills properly and efficiently. This idea is based on the interactive model (Rumelhart, 1977; Stanovich, 1980),
which posits that learners need to combine both top-down and bottom-up processing with a piece of text for comprehension. In top-down processing, learners use their prior knowledge to make predictions about a text. In bottom-up processing, learners use their linguistic knowledge to recognize linguistic elements, such as lexical items and morphological markers, in order to understand the text. In practice, learners mostly adopt a top-down approach to predict the probable theme, and from there they move to a bottom-up approach to check their understanding. Efficient learners can switch top-down and bottom-up processing according to the situation (Rumelhart, 1977; Stanovich, 1980). However, in English lessons in an EFL environment, both instruction and learning of how to switch seems to be insufficient. By way of explanation, Maeda (2017) claimed that teachers tend to focus mainly on instructing lower-level processes such as word recognition, pronunciation of words, and sentence structure analysis because many questions on entrance examinations for university mainly measure bottom-up ability rather than top-down ability. In addition, many teachers strongly believe that the lower-level (bottom-up) process is a necessary condition for moving to the higher-level (top-down) process, rather than regarding the higher-level process as a skill for compensating for weakness in the lower-level process. This focus is a tendency that occurs not only among teachers but also among researchers, who have been more interested in lower-level processes in second-language (L2) reading (Maeda, 2017). This tendency on the part of both teachers and researchers may have a relationship to a greater or lesser extent, with learners’ propensity to analyze sentences without constructing full situation models.

Previous studies have revealed that there are three steps of representation taken to form full situation models: (a) surface-level comprehension on the basis of micro-structural comprehension, including semantic and syntactic comprehension, (b) text-based comprehension on the basis of macro-structural comprehension, including the integration of prior information with present information, and (c) the construction of situation models, including the integration of information from different sentences as well as relevant information from world knowledge and the generation of inferences (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). The construction of a mental representation of the text itself is different from a representation of the situation by the text and integrated into the reader’s previous knowledge (van Dijk & Kintsch, 1983). Therefore, the textbase includes a representation of both microstructure as well as the macrostructure of the text, reflecting its local and global organization, respectively (Kintsch, 1990). As L2 learners’ cognitive resources are limited, they must be allocated to lower-level processing (Horiba, 1996, 2000; Stevenson, Schoonen, & de Glopper, 2007; Yoshida, 2003). We believe that to help learners construct situation models, teachers must de-emphasize the lower-level process and promote the higher-level process. This instruction should be influenced by the text levels and learners’ L2 reading proficiency. To what extent is teachers’ instruction, here their asking different types of questions about a text, influenced by the text levels and learners’ L2 reading proficiency?
The Effect of Questions

In the current study, the main assumption of what teachers should do in order to facilitate the switch described above is ask students questions. Ikeno (2000) said that questions about a text serve the following functions: (a) evaluating the readers’ comprehension, (b) creating an occasion for oral interaction between a teacher and students, and (c) encouraging readers’ text comprehension. In addition, Kimura (2011) said that many studies have demonstrated that questions about texts facilitate readers’ comprehension. In the current study, we investigated the effects of different types of questions on the text-based comprehension phase and construction of situation models phase. Elliot (1994) noted that most teachers agree that the most effective teaching skill is questioning. Previous studies on first-language (L1) and L2 speakers revealed that questioning had positive effects on reading comprehension (e.g., Friedman & Rickards, 1981; Ikeno, 1996; van den Broek, Tzeng, Risden, & Trabasso, 2001). Kimura (2011) summarized the effects of questioning as follows: (d) asking questions can make readers pay attention to the important elements in the text, (e) asking questions can make readers elaborate on information in the text and (f) asking questions can make poor readers use the reading strategies that good readers use. In addition, Morgan and Saxton (1991) claimed that the questions can have additional effects, including (g) helping teachers keep students actively involved in lessons, (h) encouraging students to express their ideas and thoughts openly, (i) giving students the opportunity to hear different explanations of the material by their peers, (j) allowing teachers to pace their lessons and moderate student behavior, and (k) letting teachers evaluate student learning and revise their lessons as necessary. Based on these effects, questioning is an effective teaching tool. Asking questions can help learners not only articulate their understanding of the text, but also elaborate on and analyze information. The present study investigated the relationship between different types of questions and learners’ construction of situation models.

Types of Questions

In everyday English lessons at school, teachers create various types of questions, such as fact-finding questions (FFQ), backward inference questions (BIQ), and self-reference questions (SRQ). We adopted these three types of questions in the present study because they are mainly used in many English classes.

FFQs ask about facts written in the text explicitly. In order to answer this type of question, readers only have to read and understand explicit information in the text (van Dijk & Kintsch, 1983).

BIQs ask about information that is not written directly in the text. Learners must understand both implicit and explicit information. In this process, learners generate inferences based on their background knowledge or the context (van Dijk & Kintsch, 1983). There are several types of inference questions, each of which require learners to generate several types of inferences (e.g., Graesser, Singer, & Trabasso, 1994; van Dijk & Kintsch, 1983). To create BIQs requires teachers to use a higher-level technique because they are not
simply asking about the stated facts in the text, but rather formulating a delicate question that requires learners to infer unstated information based on the written information in the text. In order to create such questions, teachers themselves need to be able to read in between the lines well.

SRQs ask learners to put themselves in the protagonist’s position (e.g., “If you were John, what would you do?”). This type of question is never seen on reading tests because there can be more than one answer. Therefore, this type of question is very educational in terms of giving learners the opportunity to express their ideas and thoughts openly, as well as helping other students hear different ideas from their friends in the classroom, as we saw in Section 1.2. In addition, the Self-Referential Effect (SRE) will strengthen learners’ memory of the text (Horiuchi & Fujita, 2001). The SRE is an idea that recall rate can be improved when people are asked to remember information when it is related in some way to the self (Horiuchi & Fujita, 2001). We also assumed that SRQs may facilitate learners’ evaluation of the content and characters in the text. This effect was investigated using an FRT and evaluated by the recall rate in the experiment. We investigated how differences especially for the construction of situation models, would come out on the basis of the differences of question types.

Purpose of This Study and Research Questions

As discussed above, three types of questions are used frequently in everyday English lessons, but few studies in L2 have investigated the relationship between question types and learners’ construction of situation models. In addition, teachers treat several different levels of text and learner’s L2 reading proficiency very differently. Therefore, it would be beneficial to investigate the relationships among question types, levels of text, and learner’s L2 reading proficiency. The three research questions (RQs) investigated in this study are given below. Answering these research questions will give suggestions for everyday English classes to teachers of English, who struggle to have their students construct situation models on reading.

RQ1) To what degree do question types affect the construction of situation models?
RQ2) Do the effects of question types on constructing situation models differ according to the text difficulty?
RQ3) Do the effects of question types on constructing situation models differ according to learners’ L2 reading proficiency?

Method

Participants

A total of 89 high-school students (35 males, 54 females) participated in this study in March of 2016. They range in age from 16 to 17 years old. They are all native speakers of Japanese and none of them had experience studying abroad for one year or longer. They all
had English language education in secondary schools in Japan for more than four years. Most of them are eager to study English to pass university entrance examinations.

Materials

**L2 reading proficiency test.** The reading test in GTEC for Students was adopted in order to measure participants’ current L2 reading proficiency. The test was composed of 43 items with three categories \((M = 176.99, SD = 19.07, \text{Max} = 229, \text{Min} = 135)\). In order to answer RQ3, the participants in each of the conditions were divided into upper and lower proficiency groups. In order to confirm whether the L2 reading proficiencies of the four question conditions were homogeneous, a two-way analysis of variance (4 × 2 ANOVA) was conducted on the scores of the L2 reading proficiency test.

The results indicated that the main effect of learners’ proficiency was statistically significant, \(F(1, 86) = 163.02, p < .001\). The main effect of the types of questions, \(F(3, 86) = 1.30, p = .281\), and the interaction between question type and learners’ proficiency were not statistically significant, \(F(3, 86) = 1.04, p = .376\).

<table>
<thead>
<tr>
<th>Question types</th>
<th>Proficiency</th>
<th>(n)</th>
<th>(M)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFQ</td>
<td>Upper</td>
<td>10</td>
<td>194.27</td>
<td>9.67</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>11</td>
<td>158.20</td>
<td>13.57</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>21</td>
<td>177.09</td>
<td>21.69</td>
</tr>
<tr>
<td>BIQ</td>
<td>Upper</td>
<td>11</td>
<td>189.64</td>
<td>8.87</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>12</td>
<td>164.23</td>
<td>10.50</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>23</td>
<td>175.88</td>
<td>16.09</td>
</tr>
<tr>
<td>SRQ</td>
<td>Upper</td>
<td>9</td>
<td>184.55</td>
<td>17.08</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>12</td>
<td>157.73</td>
<td>7.42</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>21</td>
<td>171.13</td>
<td>18.80</td>
</tr>
<tr>
<td>Control</td>
<td>Upper</td>
<td>13</td>
<td>197.92</td>
<td>10.18</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>11</td>
<td>166.19</td>
<td>9.72</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>24</td>
<td>183.38</td>
<td>18.87</td>
</tr>
</tbody>
</table>

*Note.* Maximum possible score is 320. FFQ = fact-finding questions, BIQ = backward inference questions, SRQ = self-reference questions.

**Reading passages.** In order to answer RQ2, two short passages with different difficulty levels (easy and difficult text) were adopted from the STEP EIKEN Test (STEP, 2009) in Practical English Proficiency (see Appendix A). These tests were designed for Japanese learners and the familiar topics about learners’ real-life situation were appropriate; the sentences in the tests were manageable (i.e., not too taxing on working memory capacity). As
mentioned in the previous section, L2 learners’ cognitive resources are limited, and memory while processing in reading plays an important role. The total number of words was 131 words for the easy text and 314 for the difficult text. Microsoft Word 2010’s readability measurement tools measured the Flesch–Kincaid grade levels (FKGL) for the passages at 6.2 and 9.7, respectively, and Flesch reading ease for the passages at 71.3 and 60.2, respectively.

**Question conditions.** In order to answer RQ1, three experimental conditions and a control condition were set up according to the types of questions in the reading texts. Participants were randomly divided into one of the following conditions: (a) FFQ, (b) BIQ, (c) SRQ, and (d) control.

In the FFQ condition, questions require readers to look for the facts stated in the text; the answers were presented explicitly in the texts. This type of question demanded the lowest cognitive load of readers because they could answer questions by simply reading and comprehending the texts literally.

In the BIQ condition, typical questions were *why*-questions, and the participants were required to generate causal inferences to build coherence between the focal sentences and prior sentences. This integration should strengthen coherence, which should in turn strengthen memorization (Wassenburg, Beker, van den Broek, & van der Schoot, 2015).

In the SRQ condition, learners are asked to put themselves in the protagonist’s position (e.g. *If you were John, what would you do?*). This type of question is never seen on reading tests because there is more than one answer. However, this type of question is often seen in everyday English classes.

In the control group, learners read the passage without reading questions. The purpose of the present study was to investigate how questions affect the construction of situation models. Therefore, all questions were asked and answered in Japanese in order to avoid ambiguity in understanding the question itself while reading. In addition, we attempted to avoid participants’ remembering only the forms of expression even though they had little comprehension through questioning and answering.

**Procedure**

The participants were handed a booklet containing the reading texts, the FRT, and the SRT for the two texts. Participants were not allowed to skip any pages or go back to the previous pages. They were allowed to read the texts and perform the tasks at their own pace. The order of the two texts was counterbalanced in order to avoid any sequence effect on recall.

The participants read the texts and answered two questions: one in the former part and the other in the latter part of the passages, except the control condition. The types of questions differed according to the conditions (see Appendix A). The passages were presented by sentence on one page in order for the participants not to see the next sentences and the questions were presented on the following pages. The questions were written in Japanese and the participants were required to answer in Japanese, as mentioned above. When readers
construct situation models while reading, the language forms that they remember will be a
cue to answer the questions (Kintsch, Welsch, Schmalhofer, & Zimny, 1990). The recall
answer sheet was on the following page after the passages. Therefore, the participants did not
know they would be required to participate in the FRT after reading. They were also required
to participate in the SRT after the recall task. The participants were not permitted to reread
the text while performing these tasks.

Scoring and Data Analysis

We adopted the FRT and SRT in order to measure the learners’ construction of situation
models. They are classified into off-line measurements. We adopted these measurements
because we focused on the construction of situation models rather than the revision of the
situation models in the present study. Muramoto (2000) claimed that the construction of
situation models dictates how much inference learners can generate. Once a reader constructs
a situation model representation of the text, he or she can remember the information
contained in the text better (Kintsch et al., 1990). That is why, in the present study, the
construction of situation models was measured by the off-line measures.

Quantitative analysis of recall production. First, the two different difficulties of text
were divided into a set of idea units (IUs) based on Ikeno (1996). This division was
conducted by two raters, and the agreement of division between them was 88.00% for the
easy text and 82.35% for the difficult text. Disagreements were resolved through discussion.
The total number of IUs was 25 and 51 for easy and difficult, respectively. The FRT data was
analyzed with arcsine transformation due to the difference in total numbers of IUs accordin
g to the text difficulty (Maeda, 2017). The FRT data was scored on the basis of this IU. Thirty
percent of the FRT data were randomly selected and scored by the two raters individually.
The agreement between the two raters was 96.36% for the easy text, and 96.05% for the
difficult text. Disagreements were resolved through discussion, and the remaining data were
scored by one of the raters independently.

Qualitative analysis of recall production. Qualitative, as well as quantitative, analysis
of recall production was conducted. The purpose of this analysis was to reveal how much
inferential information and how much evaluative information were included. The units in the
FRT data that were not regarded as included units in the original text were analyzed. As for
evaluative information, we investigated the participants’ evaluation of the contents and
protagonists in the text. First, these units were classified into three categories: (a) inferential
information, (b) evaluative information, and (c) error information. Thirty percent of the FRT
data were parsed into units by the two raters individually. The inter-rater agreement was
86.00%. Disagreements were resolved through discussion, and the remaining data were
parsed into units by one rater independently.

Sentence recognition task (SRT). Following Muramoto (2000), the target sentences
for the SRT were created for the two texts (see Appendix B). Each text was paired with three
types of SRT targets: (a) an explicit sentence, which describes the facts directly in the text;
(b) an inference sentence, which describes the event indirectly; and (c) an inconsistent sentence, which does not mention or suggest the event in the text. All of these targets were written in Japanese to ensure that participants’ surface text memory about language forms, such as the expression or structure of word or sentence, did not have an effect on recognition judgment. Table 2 shows an example of three target sentences for the SRT. Each target was presented with a four-point scale of recognition judgment confidence (1 = low, 2 = relatively low, 3 = relatively high, and 4 = high). This confidence scale allows for a finer-grained analysis of the participants’ recognition data.

Table 2

Examples of the Three Types of Target Sentences for the SRT

<table>
<thead>
<tr>
<th>Type</th>
<th>Japanese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit</td>
<td>ケンジの祖母は沖縄に住んでいる</td>
<td>Kenji's grandmother lives in Okinawa.</td>
</tr>
<tr>
<td>Inference</td>
<td>祖母はケンジに何かお返しをしたいと思った</td>
<td>Kenji's grandmother wanted to give something in return.</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>ケンジは祖母と一緒に暮らしたいと思っている</td>
<td>Kenji wants to live with his grandmother.</td>
</tr>
</tbody>
</table>

Note. All the target sentences were presented only in Japanese.

The SRT data was analyzed to investigate the generation of inferences. Following Muramoto (2000), the recognition ratings were calculated from the participants’ responses to SRT Targets and their confidence levels, as shown in Table 3.

Table 3

Recognition Rating Calculated From the Response and Confidence Level of the SRT

<table>
<thead>
<tr>
<th>Response</th>
<th>Confidence level</th>
<th>Recognition ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (The sentence was written in the text)</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>No (The sentence was not written in the text)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

If readers activated target inferences and then encoded the inferences as a part of text memory, inferential sentences were likely to be falsely recognized as written in the text (perhaps with a high confidence level) in spite of directly unstated events, resulting in higher cognition ratings.
Results and Discussion

Free Recall Task (FRT)

The FRT production was analyzed in order to check the text-based comprehension. The percentage scores were used in the statistical analyses. The FRT scores were analyzed with arcsine transformation due to the difference in total numbers of IUs according to the text difficulty. The means for recall production are presented in Table 4. A three-factor mixed ANOVA (2 × 2 × 4 mixed ANOVA) was conducted; it included the within-participants variable of text levels (easy, difficult), and the between-participants variables of learners’ L2 reading proficiency (upper, lower), and question types (FFQ, BIQ, SRQ, and control conditions). The main effects of readers’ proficiency, $F(1, 81) = 1.90, p = .177, \eta^2 = .01$, and question type, $F(1, 81) = 0.53, p = .661, \eta^2 = .01$, were not statistically significant. However, the main effect of text difficulty was statistically significant, $F(1, 81) = 375.88, p < .001, \eta^2 = .45$.

Table 4

The Descriptive Statistics of Recall Rates Normalized by Arcsine Transformation

<table>
<thead>
<tr>
<th>Question types</th>
<th>Proficiency</th>
<th>Easy text</th>
<th></th>
<th>Difficult text</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>FFQ</td>
<td>Upper</td>
<td>10</td>
<td>.41</td>
<td>.19</td>
<td>.13</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>11</td>
<td>.47</td>
<td>.17</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>21</td>
<td>.44</td>
<td>.18</td>
<td>.14</td>
</tr>
<tr>
<td>BIQ</td>
<td>Upper</td>
<td>11</td>
<td>.39</td>
<td>.12</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>12</td>
<td>.42</td>
<td>.20</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>23</td>
<td>.41</td>
<td>.17</td>
<td>.21</td>
</tr>
<tr>
<td>SRQ</td>
<td>Upper</td>
<td>9</td>
<td>.47</td>
<td>.20</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>12</td>
<td>.43</td>
<td>.19</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>21</td>
<td>.45</td>
<td>.20</td>
<td>.20</td>
</tr>
<tr>
<td>Control</td>
<td>Upper</td>
<td>13</td>
<td>.36</td>
<td>.15</td>
<td>.13</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>11</td>
<td>.54</td>
<td>.16</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>24</td>
<td>.46</td>
<td>.17</td>
<td>.15</td>
</tr>
</tbody>
</table>

Note. FFQ = fact-finding questions, BIQ = backward inference questions, SRQ = self-reference questions.

The interactions between learners’ L2 reading proficiency and question type, $F(2, 81) = 0.75, p = .526, \eta^2 = .01$, and learners’ L2 reading proficiency and text difficulty, $F(1, 81) = 0.98, p = .325, \eta^2 = .00$, were not statistically significant, but the interaction between question type and text difficulty was statistically significant, $F(3, 81) = 4.01, p = .010, \eta^2 = .01$. More importantly, the interaction among readers’ proficiency, text difficulty, and question type was...
marginally significant, $F(3, 81) = 2.23, p = .091, \eta^2 = .01$. The main effect of text difficulty and interaction between learners’ L2 reading proficiency and question type were considered to be influenced by this two-way interaction.

The simple main effect of readers’ proficiency indicated that the recall rate of the lower-proficiency group was higher than that of the higher-proficiency group in control condition in easy text, $F(1, 81) = 5.72, p = .019$. However, no differences were observed according to learners’ L2 reading proficiency when FFQs were given in the easy text: $F(1, 81) = 0.45, p = .505$, BIQ: $F(1, 81) = 0.16, p = .689$, SRQ: $F(1, 81) = 0.33, p = .569$. Moreover, no differences were observed in the difficult text regardless of the type of questions: FFQ: $F(1, 81) = 0.22, p = .614$, BIQ: $F(1, 81) = 0.01, p = .939$, SRQ: $F(1, 81) = 0.59, p = .444$, Control: $F(1, 81) = 0.51, p = .476$. The simple main effect of text difficulty indicated that the recall rate was higher in the easy text regardless of learners’ L2 reading proficiency or question type. On the other hand, the simple main effect of question type was not observed.

**Qualitative Analysis of Recall Production**

As stated in Data Analysis, both the proportion of overall recall and the contents of recall were analyzed with the aim of investigating how much inferential information and evaluative information were included in the recall data, as well as what types of questions facilitated generation of such information. First, the number of IUs regarded as not included in the original text was counted. These units were classified into three categories: (a) inferential information, (b) evaluative information, and (c) error information. Table 5 shows the number of inferential units and evaluative units in participants’ recall data.

<table>
<thead>
<tr>
<th>Question types</th>
<th>Easy text</th>
<th>Difficult text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inferential</td>
<td>Evaluative</td>
</tr>
<tr>
<td>FFQ</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>BIQ</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>SRQ</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Control</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note. FFQ = fact-finding questions, BIQ = backward inference questions, SRQ = self-reference questions.*

Only a small number of inferential units and evaluative units were recalled in each text level. Therefore, a statistical analysis was not performed. As indicated in Table 5, the number of inferential units depended on the text levels. The inferential units in BIQ condition were especially produced compared to other conditions. Some evaluative units were produced only
in the SRQ condition in the difficult text level. The number of inferential units in recall data by BIQ in the easy text and the number of evaluative units in recall data by SRQ in the difficult text were significant.

Table 6 shows the contents of inferential and evaluative units in recall. It seems that the participants generated the emotional inferences because of the actions of characters in the story (i.e., *His grandmother must have been happy* and *People are happy to learn how much calories food contained*).

<table>
<thead>
<tr>
<th>Text difficulty</th>
<th>Production types</th>
<th>Protocol</th>
</tr>
</thead>
</table>
| Easy            | Inference        | - *His grandmother must have been happy.*  
|                 |                   | - *His grandmother made a scrapbook in return.*  
|                 |                   | - *His grandmother made a scrapbook by referring to Kenji’s idea.*  
| Difficult       | Inference        | - *A new law was made so that people can know how many calories are in each dish.*  
|                 |                   | - *The number of restaurants that show calories has increased.*  
|                 |                   | - *People are happy to learn how much calories food contained.*  
| Evaluation      |                   | - *Both restaurants and citizens seem to benefit.*  
|                 |                   | - *The new law seems to work well.*  
|                 |                   | - *I guess that food with low calories will sell well.*  
|                 |                   | - *I believe that this system will make us healthier.*  
|                 |                   | - *If I were a customer in restaurants, I would be pleased to be able to see the calories.*  

*Note.* The original recall descriptions by the participants were written in Japanese.

It additionally seems that the participants generated bridging inferences to integrate an event with another event (i.e., *His grandmother made a scrapbook in return* and *The number of restaurants that show calories has increased*). These statements were not explicitly written in the text, but they could be inferred from the facts written explicitly in the text. In the evaluative units, participants gave statements of impressions (i.e., *I believe that this system will make us healthier*). In addition, the statements of analysis could be seen (i.e., *Both restaurants and citizens seem to benefit*), and the standpoint in the story could be seen (i.e., *If I were a customer in restaurants, I would be pleased to be able to see calories*). The evaluative questions were considered to have facilitated these descriptions in recall.

As shown in Table 6, we found the effectiveness of SRQs in the difficult text. It is assumed that learners who were given SRQs read the text, made the linkage with the events
and finally made an evaluation of the text. This did not happen in the easy text and other question conditions in the difficult text. This result suggested that SRQs would give the opportunity to have learners think deeply about the text.

Sentence Recognition Task (SRT)

The mean recognition ratings for the SRT were analyzed in order to check the construction of situation models by means of SRT (Muramoto, 2000). Specifically, the mean recognition ratings between inference sentences and inconsistent sentences were compared and the difference of rating scale value of sentence recognition (inference > inconsistent) revealed the generation of inference. The mean recognition ratings for each sentence type and text condition are reported in Table 7 for the easy text and Table 8 for the difficult text. To examine the generation of inference, the data was analyzed with a 95% confidence interval (Plonsky, 2015) instead of ANOVA due to the large and complicating number of factors, including sentence type in SRT, readers’ proficiency, text difficulty, and question types.

Table 7
Mean Recognition Ratings for SRT (Easy Text)

<table>
<thead>
<tr>
<th>Question types</th>
<th>Proficiency</th>
<th>Inference</th>
<th>Inconsistent</th>
<th>M diff</th>
<th>95% CI</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>FFQ</td>
<td>Upper</td>
<td>3.08</td>
<td>1.05</td>
<td>3.10</td>
<td>1.17</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>2.84</td>
<td>1.29</td>
<td>2.89</td>
<td>1.13</td>
<td>-0.05</td>
</tr>
<tr>
<td>BIQ</td>
<td>Upper</td>
<td>2.75</td>
<td>0.99</td>
<td>2.88</td>
<td>1.11</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>3.16</td>
<td>1.35</td>
<td>2.91</td>
<td>1.31</td>
<td>0.25</td>
</tr>
<tr>
<td>SRQ</td>
<td>Upper</td>
<td>3.09</td>
<td>1.13</td>
<td>2.60</td>
<td>1.29</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>3.34</td>
<td>1.18</td>
<td>2.77</td>
<td>1.18</td>
<td>0.57</td>
</tr>
<tr>
<td>Control</td>
<td>Upper</td>
<td>3.54</td>
<td>1.31</td>
<td>2.79</td>
<td>1.27</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>3.02</td>
<td>1.02</td>
<td>2.91</td>
<td>1.10</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Note: The ratings ranged from 0 to 6.
Table 8

Mean Recognition Ratings for SRT (Difficult Text)

<table>
<thead>
<tr>
<th>Question types</th>
<th>Proficiency</th>
<th>Inference</th>
<th>Inconsistent</th>
<th>M_diff</th>
<th>95% CI</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>FFQ</td>
<td>Upper</td>
<td>2.92</td>
<td>1.18</td>
<td>3.06</td>
<td>0.89</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>2.59</td>
<td>0.97</td>
<td>2.41</td>
<td>1.06</td>
<td>0.18</td>
</tr>
<tr>
<td>BIQ</td>
<td>Upper</td>
<td>2.68</td>
<td>1.10</td>
<td>2.72</td>
<td>1.20</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>2.93</td>
<td>1.30</td>
<td>2.45</td>
<td>1.25</td>
<td>0.48</td>
</tr>
<tr>
<td>SRQ</td>
<td>Upper</td>
<td>3.19</td>
<td>1.05</td>
<td>2.74</td>
<td>1.09</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>3.14</td>
<td>1.14</td>
<td>2.98</td>
<td>1.30</td>
<td>0.17</td>
</tr>
<tr>
<td>Control</td>
<td>Upper</td>
<td>2.86</td>
<td>1.39</td>
<td>2.36</td>
<td>1.37</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>2.93</td>
<td>1.12</td>
<td>2.60</td>
<td>1.24</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Note: The ratings ranged from 0 to 6.

For the upper group, the ratings for inference sentences were higher than those for inconsistent sentences in the easy text in the control condition, $M_{diff} = 0.75$, 95% CI [0.19, 1.31], Cohen’s $d = 0.58$. However, no difference was observed in the difficult text, $M_{diff} = 0.50$, 95% CI [-0.12, 1.12], Cohen’s $d = 0.36$. This result confirmed that participants in the upper group could not generate inferences in the difficult text without questioning, but they could generate inferences in the easy text.

However, the ratings for inference sentences were higher than that of inconsistent sentences not only in the easy text, $M_{diff} = 0.49$, 95% CI [0.05, 0.93], Cohen’s $d = 0.40$, but also in the difficult text, $M_{diff} = 0.44$, 95% CI [0.07, 0.82], Cohen’s $d = 0.41$, when participants in the upper group were given SRQs. This result indicated that SRQs activate the generation of inferences in the difficult text.

On the other hand, no difference was observed between the ratings for inference sentences and for inconsistent sentences, even when FFQs or BIQs were given: higher-proficiency group × FFQ × difficult text, $M_{diff} = -0.14$, 95% CI [-0.57, 0.29], Cohen’s $d = -0.13$; higher-proficiency group × BIQ × difficult text, $M_{diff} = -0.04$, 95% CI [-0.42, 0.34], Cohen’s $d = -0.04$. Moreover, no difference was observed between the ratings for inference sentences and for inconsistent sentences even in the easy text: higher-proficiency group × FFQ × easy text, $M_{diff} = -0.03$, 95% CI [-0.38, 0.33], Cohen’s $d = -0.02$; higher-proficiency group × BIQ × easy text, $M_{diff} = -0.14$, 95% CI [-0.47, 0.20], Cohen’s $d = -0.13$. Thus, there is a possibility that FFQs and BIQs do not promote the generation of inferences in the difficult text, and, on the contrary, interfere with the generation of inferences in the easy text.

As for the lower-proficiency group, no difference was observed between the ratings for inference sentences and for inconsistent sentences without questioning both in the easy text, $M_{diff} = 0.11$, 95% CI [-0.16, 0.39], Cohen’s $d = 0.11$ and in the difficult text, $M_{diff} = 0.33$, 95% CI [-0.08, 0.73], Cohen’s $d = 0.28$. This result did not change when FFQs and BIQs were
given: lower-proficiency group × FFQ × easy text, $M_{diff} = -0.05$, 95% CI [-0.59, 0.50], Cohen’s $d = -0.04$; lower-proficiency group × FFQ × difficult text, $M_{diff} = 0.18$, 95% CI [-0.23, 0.59], Cohen’s $d = 0.18$; lower-proficiency group × BIQ × easy text, $M_{diff} = 0.25$, 95% CI [-0.37, 0.87], Cohen’s $d = 0.19$; lower-proficiency group × BIQ × difficult text, $M_{diff} = 0.48$, 95% CI [-0.09, 1.05], Cohen’s $d = 0.38$. However, although no difference was observed in the difficult text, $M_{diff} = 0.17$, 95% CI [-0.31, 0.64], Cohen’s $d = 0.14$, a difference was observed in the easy text, $M_{diff} = 0.57$, 95% CI [0.07, 1.07], Cohen’s $d = 0.48$ when an SRQ was given. Thus, inferences were not generated without questioning in the lower-proficiency group, but SRQs promoted the generation of inferences when participants read the easy text.

### Conclusion

The present study has revealed the following three relationships: (a) the effects of questions on the construction of situation models, (b) the relationship between question type and text difficulty, and (c) the relationship between question type and reader proficiency level. First, the difference in question types produced different effects on reading texts. Questions had more of a positive effect in the difficult text than in the easy text. These results accord with the results in Kimura (2011). We found that the participants in the higher-proficiency group were able to generate inferences in the easy text even if questions were not given, but they were not able to do so in the difficult text. On the other hand, the participants in the lower-proficiency group were not able to generate inferences even in the easy text, but rather only in the case in which an SRQ was given.

Second, we found that question type and text difficulty had a strong relationship. SRQs facilitated the construction of situation models in the difficult text in the higher-proficiency group. On the other hand, FFQs and BIQs intervened in the process of constructing situation models in the easy text in the higher-proficiency group. In the lower-proficiency group, the participants did not generate inferences without questioning, as we mentioned above, in either the easy or difficult text; only SRQs facilitated generating inferences in the easy text.

Third, we found that there was a strong relationship between question types and readers’ proficiencies. When an SRQ was given, no difference in terms of generating inferences was observed, regardless of readers’ proficiencies. More importantly, FFQs and BIQs sometimes intervened in the generation of inferences in the easy text in the higher-proficiency group, meaning that unnecessary questions might prevent readers from reading texts smoothly. This finding was consistent with results in Maeda (2017).

The limitation of this study is related to the way of analysis. We classified the participants into upper and lower levels according to their L2 reading proficiency. We had thought that the differences of each category would be clearer by doing so, but, as a result, the number of participants in one category was a little bit small. Other ways of analysis such as regression analysis, may be used to predict a continuous dependent variable from a number of independent variables in future studies.

In the present study, the results revealed that teachers would do well to take readers’ L2
reading proficiency and text difficulty into consideration when creating lessons. As we mentioned above, Kimura (2011) insisted that questions were particularly effective in allowing students to comprehend difficult passages. Also, Maeda (2017) reported that there is a possibility that an excessive number of questions, regardless of readers’ proficiency with an easy text, would interfere with comprehension and the construction of situation models. Therefore, teachers must keep in mind that questions can sometimes produce negative outcomes in readers’ comprehension and construction of situation models according to text level; these outcomes are most likely to occur when teachers do not tailor questions to student proficiency levels. We conclude that types of question, text level, and learner’s L2 reading proficiency interact with each other; it is the teachers’ consideration of these elements in reading instruction that will decide readers’ success in constructing situation models.

Acknowledgement
I would like to thank Professor Yuji USHIRO and all my colleagues of his seminar for their beneficial comments on this paper and their kind and sincere cooperation on this study.

References


Appendices

Appendix A
Reading Materials and Questions Used in This Study

Instruction: Read each sentence and answer the following question.

Reading Text for Easy Text
Kenji lives in Boston with his mother and father, but his grandmother lives in Okinawa. Last April, Kenji’s grandmother had her 70th birthday. Kenji wanted to send her a special gift. He decided to make a scrapbook for her. After buying a notebook, he gathered some photographs that he had taken that year. Then he glued the photographs in the notebook and wrote something below each of them.

(FFQ) 何を集めましたか？
(BIQ) なぜケンジはその年にとった写真を集めたのですか？
(SRQ) もしあなたがケンジなら祖母に何をしてあげますか？

When Kenji called his grandmother on her birthday, she thanked him for the scrapbook. She said that it had given her an idea. A month later, Kenji received something in the mail from Japan. His grandmother had made a scrapbook with copies of photographs from when she was young. Kenji enjoyed seeing what his grandmother’s life was like when she was a girl.

(FFQ) 祖母は何を作りましたか？
(BIQ) なぜケンジの祖母はスクラップブックを作ったのですか？
(SRQ) もしあなたがケンジの祖母ならばケンジに何をしてあげましたか？
Reading Passage for Difficult Text

Most people know that fast food, such as hamburgers and fried chicken, usually contains a lot of fat and calories, and that if you eat too much of it, you are likely to put on weight. But do people really know which items on restaurant menus are the most fattening? In fact, research has shown that sometimes even experts cannot tell which foods have the most calories. Therefore, some people have been arguing that there should be laws to make restaurants tell customers exactly how many calories are contained in each item they sell. In 2008, New York became the first city in the United States to pass a law making restaurant chains show customers the calorie content of everything on their menus. Many restaurants had opposed the idea. This was partly because they were worried that customers might stop buying items with many calories in them.

Researchers have tried to find out whether this will actually happen, but the evidence is unclear. Some surveys suggest that customers may choose items with fewer calories, but others indicate that the new law is unlikely to have much long-term effect on people’s choices. Some restaurant chains, however, see the new law as a good opportunity. They have changed their menus to attract more customers and reduce costs. One leading chain of coffee shops, for example, has changed the milk it uses to low-fat milk. This lowers the calorie content of its drinks and costs less, leading to higher profit. Another well-known fast-food chain has reduced the size of its items and gotten rid of ones that contain a lot of calories. As a result, it has cut its costs and at the same time attracted new customers with its healthier menu. In fact, the chain has decided to start showing the number of calories in its products even in cities where restaurants are not required to.

(FFQ) ニューヨークは何において最初の都市となりましたか？
(BIQ) なぜニューヨークでは新しい法律が作られたのですか？
(SRQ) もしかたがレストランの経営者だったら新しい法律はどう思いますか？
Appendix B
Sentence Recognition Task

Instruction: Read the text and choose Yes (the statement was written in the text) or No (the statement was not written in the text). In addition, you should mark “confidence level” (from 1 to 4).

Easy Text

<明示文>
・ケンジの祖母は沖縄に住んでいる
・ケンジの祖母は昨年の4月に70歳になった
・ケンジは祖母の誕生日に電話をかけた
・ケンジは祖母からメールで何かを受け取った

<推論文>
・ケンジと祖母はあまり会うことができない
・ケンジは祖母の誕生日に祖母を喜ばせようと思った
・祖母はケンジからのプレゼントをとても喜んだ
・祖母はケンジに何か返しをしたいと思った

<不適合文>
・ケンジは祖母と一緒に暮らしたいと思っている
・ケンジは祖母に誕生日プレゼントを買ってあげた
・ケンジと祖母はよくメールでやり取りしている
・ケンジは祖母の声を聞いて懐かしく思った

Difficult Text

<明示文>
・ファストフードには高いカロリーが含まれている
・人々はレストランのどのメニューが高カロリーか分からない
・あるレストランチェーンは新しい法律をチャンスととらえた
・あるレストランチェーンはメニューを新しく変えた

<推論文>
・新しい法律ができることで人々は高カロリーのメニューを避けることができた
・新しい法律ができるまではカロリー表示に対する人々の希望とレストラン側の希望は異なっていた
・新しい法律ができてもお客は冷静に食べ物を選択した
・新しい法律は結果として一部に良い効果をもたらした
＜不適合文＞
・専門家ならカロリーがどれぐらい含まれているか分かっている
・レストランのメニューのほとんどが高カロリーであった
・新しい法律は人々の選択に長期的な影響を及ぼすと考えられていた
・ある人気のコーヒーチェーンは低脂肪の牛乳に変えてコストが余計にかかるようになった