The Interplay of Text Cohesion and L2 Reading Proficiency in Different Levels of Text Comprehension Among EFL Readers

Masaya HOSODA
Graduate School, University of Tsukuba / The Japan Society for the Promotion of Science

Abstract

The present study explored how the interplay of text cohesion and L2 reading proficiency affects EFL readers’ text comprehension at different levels (i.e., textbase, situation models). In the experiment, a total of 100 Japanese university students read a low- or a high-cohesion expository text and performed 3 comprehension tests: a free recall test (FRT) assessing memory of explicit ideas in the text, a causal-question test (CQT) targeting understanding of relations between text events, and a problem-solving transfer test (PSTT) measuring learning from the text. Results showed that text cohesion facilitated memory of explicit text ideas regardless of readers’ L2 reading proficiency level. In contrast, a Proficiency × Cohesion interaction was found for understanding of relations in the text: high-proficiency readers benefited from the low-cohesion text, whereas low-proficiency readers performed better when reading the high-cohesion text. Finally, a proficiency effect, but no cohesion effect was observed for learning from the text. The findings suggest that text cohesion affects textbase-level comprehension more than situation-model-level comprehension, whereas L2 reading proficiency consistently influences EFL readers’ comprehension. Based on these findings, this study suggests pedagogical implications tailored to specific levels of comprehension and characteristics of the reader and text.

1. Introduction

Reading is one of the most powerful tools through which learners acquire much of their knowledge and understanding of subject matter. Successful text comprehension is thus essential in educational settings, particularly in university and college, where English for specific purposes (ESP) education, in which students are expected to develop their expertise by reading English texts in their area of study, is required. However, it is often difficult for English as a foreign language (EFL) learners to achieve text comprehension deep enough to allow learning from the text. This difficulty is related to the fact that successful text comprehension is complicated by both reader and text variables associated with EFL reading (e.g., Gilliam, Magliano, Millis, Levinstein, & Boonthum, 2007; McCrudden, Schraw, & Lehman, 2009; McNamara, Kintsch, Songer, & Kintsch, 1996; Yano, Long, & Ross, 1994). For instance, texts do not always explicitly provide all
the information or linguistic cues necessary for coherent comprehension (e.g., McNamara et al., 1996; Ozuru, Briner, Best, & McNamara, 2010). This is particularly evident in the case of authentic expository texts, which are instrumental in ESP education, because the authors of those texts usually assume a certain amount of relevant knowledge on the part of readers. When explicit text information is insufficient, readers have to infer necessary ideas to fill in conceptual gaps in the text (Hosoda, 2015; Kintsch, 1994; McNamara et al., 1996; McNamara & Kintsch, 1996; Ozuru et al., 2010). Nevertheless, such inferential processing is particularly difficult for EFL and L2 readers because their basic reading skills (e.g., word recognition, syntactic parsing) are much less automated and thus demand more of their limited cognitive resources, as compared to L1 readers (Horiba, 2000; Hosoda, 2014, 2015). Although how well readers comprehend texts is assumed to depend on such reader-text interplay, little research has addressed this issue among EFL readers by simultaneously and systematically considering relevant reader and text variables. The general purpose of the present study is therefore to explore how text and reader factors interactively affect text comprehension in EFL readers. Of primary significance, this study focused on differences between levels of text comprehension, and how multilevel comprehension in EFL readers is impacted by text cohesion and L2 reading proficiency, both of which have been shown to influence understanding of and learning from the text (e.g., Horiba, 2000; Hosoda, 2014; 2015; McNamara et al., 1996; O’Reilly & McNamara, 2007; Ozuru et al., 2010; Yoshida, 2003).

1.1 Different Levels of Text Comprehension: Memorizing Versus Learning

Reading is generally assumed to involve the construction of multilevel mental representations of the text (e.g., Kintsch, 1994, McNamara et al., 1996). The levels of representations that are particularly relevant to discourse-level comprehension are those of the textbase and situation models. The textbase represents explicit meaning as composed of individual phrases or sentences. In contrast, situation models reflect (the overall situation of) what the text is about, where text information and readers’ knowledge are integrated. Kintsch (1994) emphasized the distinguishing of these two levels, maintaining that the textbase corresponds to the “memorizing” of the text, because this level of comprehension is largely dependent on what is explicit in the text. Thus, textbase-level comprehension can be assessed by reproductive measures such as a recall test. In contrast, situation models constitute “learning” from the text, as constructing these models involves integrating text information with the reader’s knowledge in long-term memory. Hence, assessing situation-model-level comprehension requires readers to apply learned information to a new situation in order to verify that the information consists in their long-term memory. One way to tap into situation-model-level comprehension is the problem-solving transfer test (e.g., McCrudden et al., 2009; McNamara et al., 1996). In McCrudden et al.’s (2009) experiments, L1 college readers were asked to answer such questions as “How could a space station be designed so that astronauts would be less likely to develop kidney stones?” after reading a text on the formation of kidney stones during astronauts’ space
travel. As such, simply recalling an idea or two from the text cannot solve this test. Successful problem solving requires an integrated understanding of causal relations and underlying principles, allowing readers to go beyond the text in developing an answer, which is a prerequisite for measuring situation-model-level comprehension (Kintsch, 1994; McNamara & Kintsch, 1996).

In light of these views, previous studies with EFL readers are limited in that most text comprehension measures they have used are restricted in their application to within the context of the text (e.g., Horiba, 2000; Ushiro et al., 2015); it has been rare to see tasks that require readers to go beyond the text. For example, Ushiro et al. (2015) had Japanese EFL university students answer a causal question, “Why does staying in zero gravity lead to heart shrinkage?” after students read a text on that topic (see Table 1). Although this question requires understanding of relations between text events, the exact target is presumably textbase-level comprehension, because what is required is to integrate text information within the context of the text, not to apply learned information beyond the text. Given this textbase-level focus of previous EFL reading studies, little detailed information is currently available on EFL readers’ multilevel comprehension, in particular their situation-model-level understanding.

1.2 Roles of Text Cohesion in Reading

How well readers construct multilevel mental representations of texts is a function of the complex interaction between the reader and the text (e.g., Horiba, 2000; Kintsch, 1994; McNamara et al., 1996). What is explicit in a text is usually insufficient for the construction of coherent situation models. This situation can be understood in terms of a text variable called text cohesion, which means the extent to which text ideas and the relations between them are explicit (e.g., McNamara et al., 1996; O’Reilly & McNamara, 2007; Ozuru et al., 2010). Text cohesion varies depending on the ways in which adjacent sentences are connected, such as the degree of argument overlap between sentences or the presence of specific linguistic cues (e.g., connectives including therefore, however, and moreover) that support readers to link ideas across sentences. Thus, when reading high-cohesion texts, the majority of information necessary to maintain coherent understanding is provided by the text itself (see Table 1). In contrast, when reading low-cohesion texts, readers need to engage in inferential processing to fill in conceptual gaps in the text in order to build coherence across sentences. For this reason, increased text cohesion is especially helpful for readers who have difficulty with inference generation or integrative text processing in general, particularly for textbase-level comprehension (i.e., the understanding of information within the text). Loxterman, Beck, and McKeown (1994) showed that L1 readers with low reading proficiency recalled more information from high- than from low-cohesion texts. In a similar vein, Yano et al. (1994) reported that elaborating textual descriptions to clarify implicit information facilitated comprehension in L2 readers, as measured by questions that assessed integration of information within the text.
When people first considered space travel, they did not know how the zero gravity of space would affect humans. In fact, the human body is a complex system that automatically responds to the lack of gravity. While in space, the body is not affected by gravity. Therefore, blood and water do not travel to the lower parts of the body, especially the legs. Instead, the blood and water within the body move to the upper body. Because the blood and water travel to the upper parts of the body, the body feels like the chest and head are filled with blood and water. Because of this, the heart and lungs send messages that the amount of blood and water in the upper part of the body must be reduced. As a result, space travelers do not feel thirsty, and therefore, space travelers drink less water. As body water is eliminated, their body water levels become lower than normal. When the amounts of blood and water decrease, it becomes more difficult for the human body to work normally. In addition, the decreased body water makes the heart pump less blood than normal. Therefore, the heart does not need to work as hard as it does on Earth. As a result, the heart becomes smaller. Studying the effects of space travel on humans can help us better understand many illnesses, such as high blood pressure and other heart problems.

On the other hand, prior research has also pointed out that high cohesion can be detrimental to the comprehension of readers who can engage in inferential processing by themselves when reading texts with cohesion gaps. Researchers refer to this as the reverse cohesion effect and hypothesize that high-cohesion texts provide information that is to some substantial degree redundant for such readers, which causes them to process those texts only shallowly, as compared to less cohesive texts (McNamara et al., 1996; O’Reilly & McNamara, 2007). McNamara et al. (1996) demonstrated this empirically by showing that L1 college readers with rich background knowledge perform better on tasks targeting situation models when reading low- than when reading high-cohesion texts. Using more technical texts on biology, O’Reilly and McNamara (2007) found the reverse cohesion effect for textbase questions requiring explicit text content that could be gleaned from one sentence. Thus, whether text cohesion is beneficial or detrimental is determined by readers’ propensities regarding inference generation.

### 1.3 Characteristics of L2 Reading: Why Proficiency Matters

Several recent studies have explored whether and how EFL and L2 readers make inferences during reading, and they have consistently argued that the success of L2 readers’ inference generation largely depends on those readers’ L2 reading proficiency (e.g., Horiba, 2000; Hosoda, 2014, 2015; Yoshida, 2003). The basic idea underlying this is that if L2 reading proficiency is low, readers are forced to allocate more of their limited cognitive recourses to basic reading processing (e.g., interpretation of each word and phrase), which in turn reduces resources available for discourse-level processing, including inference generation. Hosoda (2014) compared Japanese EFL university students with different L2 reading proficiency in terms of their causal inference generation from expository texts. The results showed that low-proficiency readers were
significantly inferior to high-proficiency readers in both of on-line (during-reading) and off-line (task-induced) inference generation.

At the same time, previous research has also shown that even L2 readers can make robust inferences if they possess sufficient L2 reading proficiency. Yoshida (2003) indicated that Japanese EFL university students with high L2 reading proficiency actively make elaborative inferences during narrative reading. Using expository texts, Hosoda (2015) showed further that high-proficiency EFL readers make robust causal inferences even from texts requiring inferences of unfamiliar ideas. Collectively, these studies suggest that there is a threshold of L2 reading proficiency that must be reached before L2 readers can actively engage in inferential processing.

Although the above findings are informative, little research has addressed whether and how text cohesion and L2 reading proficiency interact in EFL readers’ text comprehension. In particular, few studies with EFL readers have systematically differentiated textbase and situation-model comprehension, which is crucial to discriminate memorizing versus learning from the text. Hence, the nature of the reader-text interplay in EFL readers’ text comprehension continues to remain unclear.

Motivated by these issues, the present study aimed to gain a detailed picture of EFL readers’ different levels of text comprehension by simultaneously considering text cohesion and L2 reading proficiency. Specifically, for textbase-level comprehension, the present study examined whether and how EFL readers’ memory of explicit text information, as measured by a free recall test, is influenced by text cohesion and by L2 reading proficiency (research question: RQ1). Additionally, this study explored whether and how EFL readers’ comprehension of relations between text ideas is affected by these reader and text variables (RQ2) using a causal question test (Ushiro et al., 2015). Last, this study explored whether and how EFL readers’ situation models, as assessed by the problem-solving transfer test (McCrudden et al., 2009), are affected by text cohesion and L2 reading proficiency (RQ3), in order to attain a better understanding of learning from the text in Japanese EFL readers.

2. Method

2.1 Participants

Participants were a total of 100 Japanese university students. Their majors ranged across the humanities, education, psychology, physics, and medicine. Data from three participants were excluded because they did not complete some of the experimental procedures. The remaining 97 participants were classified into either a low- or a high-proficiency group based on their L2 reading proficiency test scores ($M = 9.98$, $SD = 5.26$, Min/Max = 3/26); 49 participants formed the high group ($M = 13.76$, $SD = 4.56$, Min/Max = 9/26), while the other 48 were the low group ($M = 5.83$, $SD = 1.68$, Min/Max = 3/8). Participants were further grouped into those who read the low-cohesion text and those who read the high-cohesion text. Thus, there were, in total, four
participant groups, crossing 2 (Proficiency) × 2 (Cohesion) conditions: (a) high-proficiency readers reading the high-cohesion text \( (n = 26) \), (b) high-proficiency readers reading the low-cohesion text \( (n = 23) \), (c) low-proficiency readers reading the high-cohesion text \( (n = 24) \), and (d) low-proficiency readers reading the high-cohesion text \( (n = 24) \). A 2 × 2 (Proficiency × Cohesion) analysis of variance (ANOVA) on L2 reading proficiency test scores confirmed that the high-proficiency readers scored better than the low-proficiency readers, \( F(1, 93) = 124.79, p < .001, \eta_p^2 = .57 \). Neither a main effect of the cohesion, \( F(1, 93) = 0.18, p = .668, \eta_p^2 = .01 \), nor the Proficiency × Cohesion interaction was significant, \( F(1, 93) = 0.00, p < .968, \eta_p^2 = .01 \).

2.2 Materials

L2 reading proficiency test. The L2 reading proficiency test aimed to examine how well readers comprehend discourse-level information, with 26 items from reading sections of the pre-first and second grades of the STEP test (Obunsha, 2005a, 2005b). The STEP test was used because this study focused on abilities related to discourse-level comprehension rather than specific grammatical or lexical knowledge, and researchers and teachers in Japan widely use the STEP test for this purpose. The reliability of the test was acceptable, Cronbach’s \( \alpha = .83 \).

Text and cohesion manipulation. The text was an expository passage discussing how zero gravity affects the human body (see Table 1). It was adapted from past expository text comprehension studies (McCrudden et al., 2009; Ushiro et al., 2015) for the following two reasons: (a) text content was assumed to not be too technical for the present participants, because this was used in Ushiro et al. (2015) who experimented with Japanese EFL readers with similar characteristics to those in this study, and (b) it has a relatively clear cause–effect structure, making it easier for text cohesion to have an effect (O’Reilly & McNamara, 2007). The difficulty of lexical items was controlled so that participants would not have much trouble with lexical-level processes; low-frequency words of level 5 or higher in the JACET 8000 List (JACET, 2003) were replaced with plainer words (level 4 or lower).

Text cohesion was manipulated by following the procedures specified in Ozuru et al. (2010). First, a low-cohesion text was created by replacing nouns with pronouns and removing connectives (e.g., however, because, and therefore). A high-cohesion text was then created by taking the original text and (a) adding connectives, (b) replacing pronouns with corresponding noun phrases, (c) adding nouns to enhance argument overlap between sentences, and (d) adding explanatory information to clarify relations between adjacent sentences (see Table 1). Two raters conducted the cohesion manipulation; when necessary, discussion was held with an English native speaker to ensure that expressions and discourse contexts were natural.

To confirm that the above manipulation altered text cohesion, key text features related to cohesion were monitored using Coh-Metrix, a computer-based tool calculating measures of text features (Graesser, McNamara, Louwerse, & Cai, 2004). As Table 2 indicates, the cohesion manipulation increased the text length by 26% and the average number of words per sentence by
3.5 words, a degree of increase commonly seen in past studies (e.g., McNamara et al., 1996). Features indicative of levels of cohesion include argument overlap between adjacent sentences, the latent semantic analysis (LSA) cosine between adjacent sentences, and the number of causal and logical connectives (Ozuru et al., 2010). As in Table 2, the cohesion manipulation successfully increased these features. In contrast, the absence of change in content word frequency implies that the manipulation did not substantially change text content itself. Last, Flesch–Kincaid grade level, a more conventional index of text readability, indicated that the high-cohesion text was more difficult than the low-cohesion text; this is to be expected, because this index is computed based on word length and sentence length, both of which were increased by the manipulation.

### Table 2 Features of Low- and High-Cohesion Texts Monitored by Coh-Metrix

<table>
<thead>
<tr>
<th>Feature</th>
<th>Low cohesion</th>
<th>High cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument overlap between adjacent sentences</td>
<td>0.77</td>
<td>1</td>
</tr>
<tr>
<td>LSA overlap between adjacent sentences</td>
<td>0.25</td>
<td>0.38</td>
</tr>
<tr>
<td>Causal connectives</td>
<td>10.64</td>
<td>42.19</td>
</tr>
<tr>
<td>Logical connectives</td>
<td>21.28</td>
<td>50.63</td>
</tr>
<tr>
<td>Content word frequency</td>
<td>2.34</td>
<td>2.32</td>
</tr>
<tr>
<td>Flesch–Kincaid grade level</td>
<td>6.72</td>
<td>8.24</td>
</tr>
<tr>
<td>Number of words per sentence</td>
<td>13.43</td>
<td>16.93</td>
</tr>
<tr>
<td>Number of words overall</td>
<td>188</td>
<td>237</td>
</tr>
</tbody>
</table>

**Comprehension tests.** Three tests were used to assess participants’ comprehension at different levels: a free recall test (FRT), a causal-question test (CQT), and a problem-solving transfer test (PSTT). Table 3 summarizes the details of comprehension targeted by each test.

### Table 3 Comprehension Tests and Targeted Representation Levels and Comprehension

<table>
<thead>
<tr>
<th>Test</th>
<th>Targeted representation level</th>
<th>Features of targeted comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRT</td>
<td>Textbase</td>
<td>Memory of explicit text ideas</td>
</tr>
<tr>
<td>CQT</td>
<td>Textbase</td>
<td>Understanding of relations between text events</td>
</tr>
<tr>
<td>PSTT</td>
<td>Situation model</td>
<td>Learning from the text</td>
</tr>
</tbody>
</table>

**FRT.** The FRT measures textbase-level comprehension, specifically, the memory of explicit ideas in the text. This test required readers to write down as much information as they can remember about the text immediately after the reading of it.

**CQT.** The CQT was intended to elicit the ability to coherently understand relations between events described in the text. The target of the CQT is supposed to be the textbase level, because required understanding consists within the context of the text (Kintsch, 1994). The CQT requires that readers explain why a certain text event happened by sequencing relevant text events in a causally appropriate order. This study’s CQT specifically asked readers to write down an explanation for why “staying in zero gravity” leads to “the heart becoming smaller” in a causally and logically coherent order. The instruction was as follows: “Why does ‘staying in zero gravity’
lead to ‘the heart becoming smaller’? Explain as much as possible to link these two events in a logically and causally correct order.”

PSTT. The PSTT was created to assess situation models, that is, learning from the text (Kintsch, 1994; McCrudden et al., 2009). This requires readers to extend what is learned from the text to a new situation in order to explain novel events or solve problems.

Questions were constructed through a pilot study in which 43 Japanese EFL university students participated. Six candidate questions were first constructed through discussion between two raters based on the following three criteria (a) correct responses require inferences from the text but (b) do not require technical knowledge that only domain experts are assumed to possess, and (c) queried situations are not restricted to within the context of the text; these criteria were created by following McNamara and Kintsch (1996), who specified requirements for situation model measures. After reading the text with high cohesion, participants were asked to answer the six candidate questions. Based on the results, two questions were excluded because they were considered to possibly impair the reliability of the test, as participants’ responses excessively varied. The remaining four questions were used in the present experiment (see Appendix A).

2.3 Procedure

The experiment consisted of immediate and delayed (a week after text reading) sessions. Note that the PSTT needed to be done in the delayed session because the intention was to assess learning from the text, not immediate memory of it.

Immediate session. The experimenter first explained the general purpose and procedures of the study. Participants were then asked to carefully read the high- or the low-cohesion text for understanding. All participants completed the reading within five minutes. The experimenter then took the text away from participants so that they could not refer to it, and the FRT was conducted. Participants were asked to write down as much of the text as possible in Japanese. The FRT was complete within 15 minutes. After that, participants performed the CQT by writing down a causal explanation of the text events in Japanese. All participants completed the CQT within 15 minutes.

Delayed session. One week after the immediate session, participants were gathered and asked to answer four questions from the PSTT without referring to the text. They were instructed to write down answers that were as detailed as possible based on what they had learned from the text (McCrudden et al., 2009). All participants completed the PSTT within 20 minutes. Finally, they took the L2 reading proficiency test with the time limit of 30 minutes.

2.4 Scoring and Data Analysis

FRT. Two raters first divided the text into 41 idea units (IUs) based on criteria in Ikeno (1996). The inter-rater agreement was 99% with disagreements resolved through discussion. Four IUs corresponding to the last paragraph were excluded from scoring to avoid any recency effect of recall. Of the remaining 37 IUs, 32 that were common in both low- and high-cohesion texts were
used for scoring. Two raters independently scored 30% of the data. A point was given if the gist of
the IU (about two-thirds of the meaning) was recalled, but scientifically incorrect responses were
not credited. The inter-rater agreement was 92%, and disagreements were resolved through
discussion. After the scoring criteria were refined, the author scored the remaining data.

**CQT.** Following Ushiro et al. (2015), participants’ answers were scored using the
*causal diagram*, which lists a series of eight text events (E1 to E8) in a cause–effect order
(see Figure 1). Among them, E1 (“Lack of gravity”) and E8 (“Heart shrinks”) were already provided on the answer sheet and thus were removed from scoring. For the remaining six events (E2 to E7), a point was given when an answer was produced in a causally correct order according to the diagram \((Min/Max = 0/6)\). As long as the order was correct, answers were credited even when intermediate events between them were absent; however, answers that were incorrect or produced in a reversed order were not credited. For example, consider an answer “when a space traveler stays in zero gravity, his body water goes up (E2), and his body feels a lot of water (E3). So the heart does not work hard (E7) and the water level then decreases (E6).” This receives three points \((E2 + E3 + E7)\); E6 does not earn a point, because it comes after E7, which should occur earlier. Two raters independently scored 30% of the data, resulting in inter-rater agreement of 95%. With disagreements resolved through discussion, the author scored the remaining data.

**PSTT.** Following Gilliam et al. (2007), answers were scored with reference to ideal answers created in advance. First, four raters independently constructed candidate answers to each of four questions, with reference to the experimental texts. Candidate answers were then parsed into individual ideas (mostly one sentence each), and the four raters had discussion to identify necessary ideas for each answer based on whether three or more of the raters included those ideas in their answers to each question. The ideal answers were finally determined by integrating the identified necessary ideas. Scoring was conducted by judging whether and how participants’ each answer correctly and coherently provided necessary ideas as identified in the ideal answer, using a four-point scale: \(0 = \text{incorrect}, 1 = \text{vague but correct}, 2 = \text{partially complete}, 3 = \text{complete}\) \((Min/Max = 0/12)\). Two raters independently scored 30% of the data, resulting in inter-rater agreement of 93%. Disagreements were resolved through discussion, and the author scored the remaining data.

For analyses, a \(2 \times 2\) (Proficiency \(\times\) Cohesion) ANOVA was separately conducted on performance on the three comprehension tests (FRT, CQT, and PSTT). Type of tests was not included as a variable because the focus of this study was on the interactive effect of cohesion and proficiency on the respective tests, not on making a direct comparison of performance across tests.
3. Results and Discussion

3.1 FRT (Research Question 1)

Table 4 presents means and standard deviations of the FRT. The ANOVA yielded a main effect of the proficiency, \(F(1, 93) = 78.89, p < .001, \eta^2_p = .46\), indicating that high-proficiency readers recalled significantly better than low-proficiency readers. Additionally, a significant main effect of the cohesion was found, \(F(1, 93) = 5.47, p = .022, \eta^2_p = .46\); participants recalled significantly better when reading the high-cohesion text, as compared to the low-cohesion text. The Proficiency \(\times\) Cohesion interaction was not significant, \(F(1, 93) = 0.03, p = .866, \eta^2_p = .01\).

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Low-cohesion</th>
<th>High-cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>High ((n = 49))</td>
<td>.36</td>
<td>.09</td>
</tr>
<tr>
<td>Low ((n = 48))</td>
<td>.14</td>
<td>.13</td>
</tr>
<tr>
<td>Total ((N = 97))</td>
<td>.25</td>
<td>.16</td>
</tr>
</tbody>
</table>

Thus, text cohesion was shown to facilitate memory of explicit text ideas, as reflected in the significant increase in recall rates for the high-cohesion text. In addition, this held true for both proficiency groups because the Proficiency \(\times\) Cohesion interaction was absent. This is in line with prior studies, such as Loxterman et al. (1994) and Yano et al. (1994), who showed that increased cohesion promotes textbase-level understanding. Presumably, because ideas and relations were made explicit in the high-cohesion text, individual text information was strongly encoded into readers’ memory. To be more specific, this cohesion effect might be to a larger extent due to increased opportunities to encounter specific text information in the high-cohesion text than to enhanced inference generation. It is assumed that the target of the FRT is memory of explicit text ideas rather than integrative or implicit understanding of the text related to inferential processing (Kintsch, 1994; McNamara et al., 1996). From this perspective, a likely explanation for the enhanced recall for the high-cohesion text is the more frequent exposure of participants in this condition to explicit text content due to the increased argument overlap between sentences, which might help readers retain and encode the text information. This possibility may also explain the fact that the cohesion effect was observed regardless of L2 reading proficiency; encoding frequently encountered explicit information would be expected to demand fewer cognitive resources than actively inferring implicit information, and therefore, participants as a whole were able to benefit from high cohesion. Despite these observations, we cannot obtain a definitive conclusion from the FRT findings alone as to whether text cohesion positively or negatively affects comprehension associated with inferential processing. To address this, the subsequent sections discuss the results of the CQT and the PSTT.
3.2 CQT (Research Question 2)

Table 5 reports means and standard deviations of the CQT. The ANOVA revealed a main effect of the proficiency, $F(1, 93) = 46.35, p < .001, \eta^2_p = .33$, indicating that high-proficiency readers performed significantly better than low-proficiency readers. Unlike on the FRT, the cohesion main effect was not significant, $F(1, 93) = 0.70, p = .405, \eta^2_p = .01$. However, these were qualified by the significant Proficiency $\times$ Cohesion interaction, $F(1, 93) = 10.41, p = .002, \eta^2_p = .10$. To interpret this, a follow-up test was conducted, and revealed that high-proficiency readers did better when reading the low-cohesion text ($M = 3.26, SD = 1.21$) than when reading the high-cohesion text ($M = 2.27, SD = 1.43$), $F(1, 93) = 8.32, p = .005, \eta^2_p = .08$, whereas the reverse was observed in low-proficiency readers with better performance on the high-cohesion text ($M = 1.46, SD = 1.06$) than the low-cohesion text ($M = 0.88, SD = 1.03$), though this remained a marginally significant trend, $F(1, 93) = 2.83, p = .096, \eta^2_p = .03$.

Table 5 Mean Scores and Standard Deviations of the CQT

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Low-cohesion</th>
<th>High-cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>High ($n = 49$)</td>
<td>3.26</td>
<td>1.21</td>
</tr>
<tr>
<td>Low ($n = 48$)</td>
<td>0.88</td>
<td>1.03</td>
</tr>
<tr>
<td>Total ($N = 97$)</td>
<td>2.04</td>
<td>1.64</td>
</tr>
</tbody>
</table>

*Note. Scores range from 0 to 6.*

The CQT results therefore exhibited the reverse cohesion effect in high-proficiency readers, who seemed to better understand relations between text events when reading the low- than the high-cohesion text. According to past L1 reading research (e.g., McNamara et al., 1996; O’Reilly & McNamara, 2007), one explanation for the reverse cohesion effect is that readers who can infer information to fill in cohesion gaps by themselves may fall into shallower processing with high-cohesion texts, as relative to low-cohesion texts, because of the redundancy of information provided in high-cohesion texts. Given this explanation and the existing finding that Japanese EFL readers with high L2 reading proficiency can generate robust inferences during reading (Hosoda, 2014, 2015; Yoshida, 2003), the high group reading the low-cohesion text might engage in inferential processing in order to involve loosely connected information to establish coherence between sentences. This might aid their understanding of relations that link text events to each other in the low-cohesion text, as compared to the high-cohesion text. This possibility seems likely given that the content and lexical items of the text were controlled so as not to be too technical or difficult for the present participants (i.e., Japanese EFL university students).

Alternatively, one may argue that the reverse cohesion effect in the high group was due to the increased amount of information in the high-cohesion text. Specifically, as indicated by Coh-Metrix (see Table 2), the high-cohesion text was longer, included more words per sentence, and was estimated by the readability test to be more difficult than the low-cohesion text. This greater intensity of information in the high-cohesion text might impose correspondingly greater
demands on readers to comprehend relevant text relations in an integrative manner. However, the absence of the reverse cohesion effect in low-proficiency readers, who presumably require more cognitive resources to perform integrative processing, runs counter to this possibility. Rather, a marginally significant trend was seen in which high cohesion served as scaffolding for low-proficiency readers to move beyond floor-level comprehension (see Table 5).

Regarding the fact that the reverse cohesion effect was not observed in the FRT, which is also assumed to assess textbase comprehension (Kintsch, 1994), an integrative feature of the CQT might be responsible. In contrast to the FRT, where simply recalling more text ideas increases performance, the CQT requires coherent linkage of a series of events across the entire text. Hence, the understanding needed on the CQT would be expected to require more extensive information than on the FRT, which might increase the prominence of inferential processing, a main cause of the reverse cohesion effect (McNamara et al., 1996). This view in turn suggests that, although remaining marginally significant, the better CQT performance in the low group for the high-cohesion text implies that text cohesion helped low-proficiency readers’ inferential processing related to integrative understanding.

3.3 PSTT (Research Question 3)

Table 6 shows means and standard deviations of the PSTT. The ANOVA showed a significant main effect of the proficiency, $F(1, 93) = 34.25, p < .001, \eta^2_p = .27$, indicating that high-proficiency readers performed significantly better than low-proficiency readers. The cohesion main effect was not significant, $F(1, 93) = 0.05, p = .827, \eta^2_p = .01$. In addition, there was a marginally significant effect of the Proficiency × Cohesion interaction, $F(1, 93) = 3.37, p = .070, \eta^2_p = .04$. However, a follow-up on this showed only that high-proficiency readers scored better than low-proficiency readers at both cohesion levels: high cohesion, $F(1, 93) = 8.32, p = .005, \eta^2_p = .08$; low cohesion, $F(1, 93) = 28.68, p < .001, \eta^2_p = .24$. On the other hand, text cohesion had no significant effect at either proficiency level (both ps > .10).

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Low-cohesion</th>
<th>High-cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>High ($n=49$)</td>
<td>6.00</td>
<td>3.30</td>
</tr>
<tr>
<td>Low ($n=48$)</td>
<td>1.79</td>
<td>1.38</td>
</tr>
<tr>
<td>Total ($N=97$)</td>
<td>3.85</td>
<td>3.27</td>
</tr>
</tbody>
</table>

Note. Scores range from 0 to 12.

Thus, the PSTT results showed that proficient readers learned more from the text than less proficient readers. However, in contrast to the other tests, there was no significant beneficial or detrimental effect of cohesion in either proficiency group. One possible reason for the absence of the cohesion effect might relate to the level of comprehension assessed by the PSTT. Correctly answering the PSTT requires extending information gained from the text to outside the context of
the text, and consequently, there might be less room for text cohesion, which varies depending on what is explicit in the text, to act a significant role. In support of this view, previous L1 reading research has often reported that the benefit of cohesion is more evident for textbase-level comprehension, such as understanding information within a sentence, than for situation-model-level comprehension, involving integration of information across distant sentences or integrating it with prior knowledge (e.g., Loxterman et al., 1994; O’Reilly & McNamara, 2007). Taking one question from the PSTT as an example—“Explain why people are advised to refrain from drinking a lot of water when their heart is weakened”—to answer this question, readers first need to learn that the functioning of the heart depends on the amount of body water, which they can inferentially understand from the text’s statement that space travelers’ hearts do not work hard in space because their body water decreases. Then, readers are further required to utilize this idea to reason that ingesting a lot of water may pose an excessive burden on the heart, because the amount of water to be circulated is increased. Thus, merely remembering explicit text ideas and relations cannot solve the PSTT. This applied character of the PSTT might make it less likely for text cohesion to have a notable effect.

This possibility may be augmented when considering the features of the cohesion manipulation employed in this study. We manipulated text cohesion mainly at a local level, by replacing pronouns with specific noun phrases, inserting connectives to link sentences, and adding some explanatory information to make meanings of adjacent sentences clearer. At a global level, however, the present manipulation hardly changed cohesion: possible global changes might include inserting additional paragraphs or providing large amounts of background information (McNamara et al., 1996). It might be thus easier for text cohesion, altered mostly by the linguistic items within the text, to impact textbase than situation-model comprehension, whether positively or negatively. This at the same time suggests that, at least in the present experimental environment, simply adding linguistic signals and explanatory information is unlikely to be sufficient to promote EFL readers’ learning from the text; they might need additional background information related to the text to support their text learning.

4. Conclusion

The present study examined EFL readers’ different levels of text comprehension in light of the interplay of text cohesion and L2 reading proficiency. Findings regarding textbase-level comprehension showed that increased cohesion enhanced memory of explicit text ideas in both low- and high-proficiency readers (RQ1). In contrast, the Proficiency × Cohesion interaction was found to affect understanding of relations between text events; high-proficiency readers benefited from low-cohesion text, whereas the opposite trend was observed in low-proficiency readers, for whom high cohesion supported an escape from the floor-level performance (RQ2). Finally, for situation-model comprehension, high-proficiency readers learned more from the text than the
low-proficiency readers, but text cohesion had no significant effect. Collectively, these findings suggest that while L2 reading proficiency consistently impacts EFL readers’ comprehension, text cohesion, manipulated by the number of linguistic cues and amount of explicit information, is more likely to influence textbase than situation-model comprehension, whether that influence is beneficial or detrimental. In this regard, this study is informative in that it proves among EFL readers that the contribution of the text variable to comprehension systematically varies and interacts with the reader variable according to differences in the levels of comprehension.

From this, we might derive some pedagogical implications tailored to targeted comprehension and characteristics of texts. First, when teachers use relatively technical texts that communicate novel information to be learned, it is advisable to increase text cohesion by adding discourse markers or explanatory information to make relations between sentences explicit. This is because when it comes to unfamiliar texts, it is primarily important to ensure that learners correctly understand what is stated in the text (i.e., to construct textbase-level comprehension; O’Reilly & McNamara, 2007), before inferring implicit ideas and actively learning from the text. In this sense, the findings from the FRT suggest that increased cohesion can help learners build basic textbase of unfamiliar texts. However, it is also important to keep in mind that increasing text cohesion is not sufficient, though may be necessary in some cases, for deep comprehension. For learners with sufficient L2 reading proficiency, providing an opportunity to make meaning from the text by themselves would be promising, as suggested by the fact that high-proficiency readers performed better on the CQT after reading low- than high-cohesion text. Giving authentic texts on their majors and instructing them to explain text content in their own words (i.e., self-explanation) would serve this line of instructions (e.g., Ozuru et al., 2010). For less proficient learners who have difficulty making inferences, on the other hand, enhancing cohesion would help them understand the text in an integrative manner. Finally, to support EFL students’ learning from the text, it seems necessary to enhance both text and reader factors. To foster students who can learn effectively from texts, it is therefore important to employ approaches that are intended to develop learners’ cognitive skills and knowledge base, plus tailoring text cohesion according to their proficiency levels. These may include implementing extensive reading activities to train learners’ ability to appropriately and accurately comprehend discourse-level information and providing sufficient amounts of background information before learners read texts.

These implications should be considered only in light of an important methodological limitation of this study: Learning outcomes from the text were assessed using only one measure. This makes it dangerous to generalize the present findings to EFL readers as a whole. Other operationalization of situation-model comprehension includes having readers categorize key text concepts based on the text (the sorting task; McNamara & Kintsch, 1996) and draw pictures of situations described in the text. To determine the generalizability of the present findings, further studies comparing the results of such measures are needed. Also meriting future research are “during-reading” processes, in particular learners’ inferential processing; findings on these
processes will complement the present study. For example, an experimenter may ask readers to think aloud their thoughts during reading sentences that are theoretically identified to be important for understanding the text (Ozuru et al., 2010). Such attempts to bridge reading processes and comprehension may offer a more comprehensive picture of EFL readers’ learning from the text.

Acknowledgements

This research was supported by a Grant-in-Aid for JSPS Fellows (No. 15J01241). I would like to thank Prof. Yuji USHIRO and my colleagues for their support. I also wish to extend my gratitude to three anonymous reviewers for their insightful comments to improve this paper.

References


Appendix A: The Problem-Solving Transfer Test Used in The Present Study

1. If you were given a command that made the lungs or heart function as if they were in a story written in the text, how would the rest of the body (internal organs, etc.) respond? Consider if the body's water level would change if the command were issued in space. (Note: This is a story in space.)

2. Heart rate is slow, please avoid excessive intake of water. Explain this, based on the relationship between the heart and the body's water levels as described in the text. (Note: This is in the story on Earth.)

3. The body, unable to perceive changes in water, contracts a disease. Even if it goes to space, the heart is not likely to change much from its size on Earth. Explain this, based on the information from the text. (Note: This is in the story in space.)

4. To prevent the heart from shrinking in space flight, what equipment or functions should be installed on the space station? (Note: This is in the story about space.)