Learning from expository text in L2 reading: Memory for causal relations and L2 reading proficiency

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

Abstract

This study explored the relation between second-language (L2) readers' memory for causal relations and their learning outcomes from expository text. Japanese students of English as a foreign language (EFL) with high and low L2 reading proficiency read an expository text. They completed a causal question and a problem-solving test as measures of memory for causal relations and learning from the text, respectively. It was found that memory for causal relations contributed to text learning in high-proficiency readers, but not in low-proficiency readers. The quantitative and qualitative analysis of causal question answers revealed that low-proficiency readers recalled fewer causal relations and made more incorrect inferences than high-proficiency ones. Additionally, low-proficiency readers tended to perform the problem solving using inappropriate causal sequences and irrelevant information. These findings suggest that low-proficiency readers struggled with processes at both textbase and situation-model levels; consequently, they failed to learn causal relations in the text as knowledge.

Keywords: L2 reading, expository text, situation models, causal relations, learning from text, L2 reading proficiency

The goal of this study was to explore the relation between second-language (L2) readers' memory for causal relations and their learning outcomes from expository text. Some expository texts communicate scientific principles or mechanisms to readers so that they can learn such information as new knowledge. Ability to learn scientific knowledge from expository text in L2 is especially important for university or college students because this ability can broaden one's expertise beyond the bounds of language and contribute to academic progress.

In the field of cognitive research on first-language (L1) reading, learning from text is construed as the construction of coherent mental representations of situations described in the text, known as *situation models* (Kintsch, 1994; McNamara, Kintsch, Songer, & Kintsch, 1996). The construction of situation models involves not only understanding explicit text elements but also inferring relations that connect pieces of text information (Kintsch, 1998). It is well known that causal relations play a central role in the processes of the situation-model construction (e.g., McCrudden, Schraw, & Lehman, 2009; Millis, Magliano, & Todaro, 2006; Mulder & Sanders,

2012).

Unlike L1 research, research is quite limited that directly investigated L2 readers' learning from text. Therefore, sufficient information is not available about how L2 students' text learning can or should be supported. The present study was intended to fill in this gap by examining the relation between memory for causal relations and L2 readers' text learning.

Situation Models and the Role of Causal Relations in Comprehension

Text comprehension is generally assumed to involve readers constructing at least three levels of mental representations: *surface code, textbase*, and *situation models* (Kintsch, 1998). Surface code is a verbatim memory of words or phrases and is less relevant to discourse processes (Mulder & Sanders, 2012). Textbase is a representation of meanings of text, comprised of textual propositions and their interconnections. Textbase is relevant to discourse processes because it includes readers' memory for relations between information in text (Kintsch, 1998; McNamara & Kintsch, 1996). Finally, situation models refer to a representation of what text is about, amounting to an amalgamation of information explicitly present in text and inferences made by readers (Kintsch, 1994; McNamara et al., 1996). As stated above, learning from text is construed as the construction of situation models. At the situation-model level, readers go beyond the text to understand implicit relations or principles by making inferences based on the previously understood text as well as prior knowledge (Kintsch, 1998; Millis et al., 2006). Thus, when situation models are successfully constructed, information conveyed by text is integrated with long-term memory knowledge and serves the applied purposes (e.g., providing explanations, solving problems).

There is ample evidence that the situation-model construction for expository text depends on readers' ability to detect and understand causal relations in text (McCrudden et al., 2009; Millis et al., 2006; Mulder, 2008; Noordman, Vonk, & Kempff, 1992). Causal relations are important for deep expository comprehension as they provide an essential framework to order information in a consistent and logical manner (León & Peñalba, 2002). Researchers have empirically demonstrated that (a) causally related information is processed faster and encoded into memory more strongly than information conveyed by other types of relations such as temporal relations (Mulder, 2008; Singer, Halldorson, Lear, & Andrusiak, 1992), (b) elaborating text by signaling causal relations enhances comprehension (Linderholm, Everson, van den Broek, Mischinski, Crittenden, & Samuels, 2000), and (c) skilled L1 readers base their standards of coherence on causality between multiple sentences (Wittwer & Ihme, 2014).

Given the importance of causal relations in the situation-model construction for expository text, readers' understanding of causal relations is linked to learning from text. However, few L2 studies have addressed this link directly. To pursue this link, I next discuss how L2 text learning is potentially impacted by L2 reading proficiency.

The Role of L2 Reading Proficiency in the Situation Model Construction

Contributions of memory for causal relations to text learning in L2 readers may depend on their L2 reading proficiency. The underlying rationale is that how well L2 readers construct inferences necessary for constructing situation models is constrained by their L2 reading skills; when L2 reading proficiency is low, readers have to prioritize basic reading processes (i.e., word decoding, syntactic parsing) over inferential processing in the allocation of their cognitive resources (Horiba, 1996; Hosoda, 2014). The limited inferential processing may then make it difficult to retain memorized causal relations in situation models (Kintsch, 1994).

Noteworthy in this regard is a study by Ushiro et al. (2015) in which Japanese university students read an expository text on how astronauts' hearts shrink in space. After reading, participants were asked to recall all the information they could remember from the text (i.e., a recall test). They then completed a causal question (called a why-question test in their study) requiring a causal explanation of the text ("Why does staying in zero gravity lead to space travelers' hearts becoming smaller?"). The researchers investigated whether the appropriateness of the causal structure in memory (whether causally important information was recalled better than less important information) led to causally explaining the text. The results showed that this was the case only for participants with high L2 reading proficiency; less proficient readers still had difficulty with the causal explanation, even when they built the appropriate causal structure in memory. The researchers attributed these findings to the readers' ability or inability to construct inferences to meaningfully connect text ideas.

Based on these findings, it is likely that readers who are relatively fluent in basic reading processes (e.g., L1 readers, advanced L2 readers) can make inferences needed to construct coherent situation models. Conversely, due to the difficulty with inferential processing, less proficient L2 readers may struggle to construct situation models, which would prevent them from learning texts' causal relations as knowledge.

Assessing Situation Models and Learning from Text

As described above, situation-model construction entails the integration of text information and readers' knowledge. To assess situation models, mainly two types of measures are available. The first is to examine inferences in readers' responses to a reproductive measure (e.g., recall, summary). McNamara and Kintsch (1996) proposed that the extent to which a reproductive measure indicates textbase or situation models is on a continuum, varying as a function of the amount of inferences included in responses; the larger the amount of inferences, the more indicative of situation models. In line with this notion, Barry and Lazarte (1998) examined the amount of three types of inferences in L2 readers' recall protocols: (a) within-text inferences: logical interpretations from text, (b) elaborative inferences: combined propositions of text ideas and readers' knowledge relevant to the text topic, and (c) incorrect inferences: propositions that are contradictory or irrelevant to the meaning of a text. They interpreted the total amount of these inferences as indicative of the richness of situation models, whereas as they did proportions of incorrect inferences (incorrect inferences / [within-text + elaborative + incorrect inferences]) as the accuracy of situation models.

An advantage of a reproductive measure is that the result is less vulnerable to semantic and syntactic constraints, compared to online measures (e.g., a lexical decision task). Still, it must be noted that this type of measure only partly assess readers' application of knowledge learned from the text. This is because a reproductive measure exclusively requires reproducing information as in the original text (e.g., recall what was written in the text).

The second measure, which overcomes the abovementioned limitation, is to ask readers to apply the learned information to a new situation to solve problems or explain phenomena. This is often referred to as a *problem-solving test* (e.g., Mautone & Mayer, 2001). In McCrudden et al. (2009), L1 college students answered such problem-solving questions as "How could a space station be designed so that astronauts would be less likely to develop kidney stones?" after reading an expository text on how kidney stones develop in space. This requires that both the mechanism underlying the development of kidney stones (e.g., astronauts' bones become weaker, calcium levels in blood get higher, and the kidney needs to filter greater amounts of calcium from the blood), and knowledge about how to improve bone strength (e.g., doing physical exercises), are available in long-term memory. In other words, successful problem solving cannot be achieved unless readers make inferences from prior knowledge, as well as from the text. As such, the problem-solving test explicitly requires readers to go beyond the text and therefore more directly targets the ability to apply what was learned from the text than a reproductive measure.

The problem, however, is that most past L2 reading studies used reproductive measures to assess situation models (e.g., Horiba, 2000; Ushiro et al., 2015). Therefore, it continues to be unresolved whether or to what extent L2 readers productively use the knowledge gained from the text in novel environments. This methodological limitation is also responsible for the limited understanding of learning from text in L2 reading.

This Study

The goal of this study was to explore the relation between memory for causal relations and learning outcomes from expository text in L2 readers. Based on the notion that inferential processing in L2 readers is constrained by L2 reading proficiency, I considered readers' L2 reading proficiency as a variable.

Readers' memory for causal relations was assessed by a causal question ("Why does staying in zero gravity lead to the heart becoming smaller?"). This why-type question is widely used to elicit readers' causal explanations (e.g., Carlson, van den Broek, McMaster, Rapp, Bohn-Gettler, Kendeou, & White, 2014; Ushiro et al., 2015). To explain text causally, readers must have an integrative understanding of a series of causal relations between pieces of text information (León & Peñalba, 2002; Millis et al., 2006). Thus, the response to such a causal question can be an indicator of the reader having understood and memorized causal relations in the text. In addition, I inspected inferences in participants' causal question answers to examine how well they constructed situation models of causal relations in the text (Barry & Lazarte, 1998).

For learning outcomes from the text, I used a problem-solving test to assess participants' application of knowledge gained from the text. Furthermore, to obtain a detailed picture of the

effect of L2 reading proficiency on problem-solving, I qualitatively examined contents of problem-solving responses and identified patterns that were distinctive of low-proficiency readers. The following research questions (RQs) were addressed in this study:

RQ1. Do contributions of L2 readers' memory for causal relations to their learning outcomes from the text differ as a function of L2 reading proficiency?
RQ2. How does L2 reading proficiency affect L2 readers' memory for causal relations?
RQ3. How does L2 reading proficiency affect L2 readers' inferences generated in the causal question?
RQ4. How are low-proficiency L2 readers characterized by contents of their problem-solving responses?

Method

Participants

Participants were 70 Japanese university students majoring in the humanities, education, medicine, physics, or psychology; 42 were female and 28 were male. Their ages ranged from 18 to 24 years old ($M = 20.01$, $SD = 3.16$). They were all native speakers of Japanese and had lived in Japan for more than 18 years. They had studied English as a foreign language (EFL) at least for six years as part of the compulsory education in Japan. None of them had experience of studying abroad for more than two weeks. Based on self-reported scores on standardized English proficiency tests (e.g., the EIKEN, TOEFL, and/or TOEIC test), their overall English proficiency was estimated to range from the CEFR A1 to B2 levels (Council of Europe, 2001) or 40 to 95 in the TOEFL iBT test (Papageorgiou, Tannenbaum, Bridgeman, & Cho, 2015).

Participants were grouped by means of a median split ($Mdn = 9$) into low- ($n = 34$, $M = 5.94$, $SD = 1.77$, $Min/Max = 3/8$) and high-proficiency groups ($n = 36$, $M = 14.67$, $SD = 5.15$, $Min/Max = 9/26$) based on their performance on an L2 reading proficiency test ($M = 10.43$, $SD = 5.85$, $Min/Max = 3/26$), $t(68) = 9.6$, $p < .01$, $d = 2.26$, This grouping generally corresponded to the self-reported proficiency levels, with participants in the high-proficiency group reporting higher proficiency than those in the low-proficiency group.

Materials

L2 reading proficiency test. The L2 reading proficiency test consisted of 26 items. These were derived from the reading section of the second ($k = 20$) and pre-first grades ($k = 6$) of the EIKEN Test in Practical English Proficiency. The EIKEN test is one of the most widely used standardized English proficiency tests among Japanese educators and researchers (e.g., Hosoda, 2014, 2015, 2016; Ushiro et al., 2015). The 26 items of the present proficiency test tap into understanding of the content of passages rather than specific lexical or grammatical knowledge. Therefore, this proficiency test is assumed to target ability to comprehend English text appropriately and efficiently. The reliability of the test was acceptable, with Cronbach's $\alpha = .81$.

Text. An expository text discussing how staying in zero gravity causes astronauts' hearts to shrink was used (presented in Appendix A). This passage was originally part of a longer passage used in McCrudden et al.'s (2009) experiment with L1 readers. Later, Ushiro et al. (2015) shortened the passage in their experiment with L2 readers. The present passage is an adapted version of Ushiro et al.'s.

Specifically, I added some statements to Ushiro et al.'s (2015) passage to make clearer the discourse flow in the later part of the passage (why the weakening of the heart functioning leads to the heart shrinkage). I also deleted statements that were not relevant to the main theme of the passage to clarify the cause-effect sequences. Further, words assumed to be unfamiliar to EFL readers were replaced with more familiar words (e.g., fluids → water). This was done with reference to the *JACET (the Japanese Association of College English Teachers) list of 8000 basic words* (Ishikawa et al., 2003). This lists English words that Japanese students are supposed to learn from elementary school to university based on frequency from level 1 (most frequent) to 7 (least frequent). Words from level 5 or higher were rephrased with more frequent words from level 4 or below. After this, a native speaker of English checked the text for the naturalness of expression and discourse.

Causal question. The causal question asked participants to causally explain why "staying in zero gravity" leads to "the heart shrinking," the series of causal relations described in the text. Expected responses included a maximum of six causal relations (CRs; CR2–CR7), which are presented in Appendix B. Note that the beginning (CR1 "Lack of Gravity") and the end (CR8 "Heart shrinks.") of the causal chain were provided as a cue on the answer sheet (a possible effect of this is discussed later).

The instruction for the causal question 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." As in the instruction, participants were told that it was important to provide as many relevant causal relations as possible, which aimed to avoid confounding with the omission of understood relations from an answer. The instruction was given on the answer sheet, in Japanese (participants' L1), and participants were also asked to answer in Japanese so that L2 writing skills would not affect the result. They were not allowed to refer to the text during the task.

Problem-solving test. The problem-solving test consisted of four questions requiring participants to use learned principles or mechanisms to solve problems or explain situations outside the original text.[1] Table 1 presents an example question and its expected response. As seen there, correct responses should not only be consistent with the text but also include inferences from the knowledge gained from the text.

Appendix C presents all the four problem-solving questions in addition to the instruction. These questions were presented in Japanese in a fixed order. Participants were asked to write down their answers in as much detail as possible in Japanese. Participants were also asked to provide reasons for each answer based on what they had learned from the text.

Table 1. *An example problem-solving question and expected response (translated from Japanese by the author)*

Question:
Imagine that you got a disease that makes you insensitive to changes in body fluids. Explain why, in this condition, your heart size would not change even if you stayed in space.
Expected response:
(If the body were insensitive to changes in fluids), the body cannot eliminate fluids even though the blood and water collect in the upper body in space. So, the amount of body fluids would not be reduced, which causes the heart to pump as strongly as on the earth. Because the heart works as usual, the heart size would not change.

Procedure

The experiment consisted of two phases. Participants were tested in groups of one to five people. In the first phase, the experimenter explained the experiment's purpose and procedures, and informed consent was obtained. Participants were asked to read the experimental text for understanding at their own pace and turn the sheet over after reading the text once. There was no time limit for reading, but all participants finished within five minutes. Subsequently, the causal question was administered. Participants completed their responses to this within 15 minutes. The second phase was conducted one week later. Participants were assembled, and the problem-solving test was administered. This was completed within 20 minutes. Finally, an L2 reading proficiency test lasting 30 minutes was administered.

Scoring

Causal question. Memory for causal relations was assessed in terms of the number of key causal relations (presented in Appendix B) recalled in the answers.[2] The answer scored 0 to 6 points depending on the number of causal relations recalled in a causally correct order. As long as the order was correct, points were given even when intermediate relations between them were absent. However, answers that were scientifically incorrect, inconsistent with the text, or produced in an incorrect order were not credited. For example, consider the answer, "When a space traveler stays in zero gravity, his body water goes up (CR2), and his body feels a lot of water (CR3). So, the heart does not work hard (CR7), and the water level then decreases (CR6)." This receives 3 points (CR2 + CR3 + CR7); CR6 does not earn a point because it comes after CR7, which should occur earlier. Two Japanese graduate students (including the author) independently scored 30% of the data, resulting in inter-rater agreement of 95%. After disagreements were resolved through discussion, the author scored the remaining data.

As for inferences in responses, two Japanese graduate students (including the author) first counted inferential ideas that they found in causal question answers through discussion. The unit of analysis was a subject-verb clause in Japanese. The found inferences were subsequently classified into the following three categories (Barry & Lazarte, 1998): (a) within-text inferences, (b) elaborative inferences, and (c) incorrect inferences. Within-text inferences indicated readers' building coherence across pieces of the text. This was often associated with adding of explanatory information to the explicit text (e.g., the underlined part of "The muscle in the heart

is reduced because it does not need to work hard"). Elaborative inferences indicated readers' embellishing mental representations with prior knowledge. This was also associated with adding relevant information to text ideas, but it did not contribute to the coherence (e.g., "While in space, an astronaut's face will swell because fluids shift to the upper body"). Incorrect inferences indicated readers' misunderstanding or use of irrelevant information to the text. This type of inference included off-topic information or ideas contradictory to what was mentioned in the text (e.g., "The heart sends greater amounts of blood than usual as the body fluids decrease in space"). The two Japanese graduate students (including the author) independently categorized 30% of the inferences, resulting in inter-rater agreement of 90%. After disagreements were resolved through discussion, the author categorized the remaining 70% of data.

Problem-solving test. Participants' responses were assessed by a scoring system developed in past expository comprehension research (Gilliam, Magliano, Millis, Levinstein, & Boonthum, 2007; Magliano, Millis, the RSAT Development Team, Levinstein, & Boonthum, 2011; Millis et al., 2006). This identified necessary information for the pre-created ideal answers for each question.[3] Responses were scored using a 4-point scale (0–3): 0 meant that the answer was incorrect; 1 meant that the answer was vague but correct on the whole; 2 meant that the answer was partially complete; 3 meant that the answer was complete. The total scores ranged 0 to 12. Two Japanese graduates (including the author) scored 30% of the data, resulting in inter-rater agreement of 93%. After disagreements were resolved through discussion, the author scored the remaining data.

For the qualitative analysis, we took a four-step procedure. First, two Japanese graduate students (including the author) separately examined problem-solving responses, and a list of response patterns that were distinctive of the low-proficiency group was constructed. Second, we had a discussion and eliminated those patterns that were not agreed on and that overlapped with other patterns. Third, proportions of participants showing the selected patterns were compared between the proficiency groups, using chi-squared tests. Finally, patterns whose *p* values were below .05 were deemed as distinctive of the low-proficiency group.

Results

The results section consists of four parts. First, I report descriptive statistics of performance on the causal question and the problem-solving test. Second, I report contributions of causal question performance to the problem-solving. Third, I report results regarding inferences in the causal question. Finally, I report distinctive patterns of problem-solving responses in the low-proficiency group.

Descriptive statistics and intergroup differences

Table 2 presents means and standard deviations of the two proficiency groups' performance on the causal question and the problem-solving test. For the causal question, the high-proficiency group performed significantly better than the low-proficiency group, $t(68) = 2.94$, $p = .004$, $d = 0.70$. Specifically, the majority of the low-proficiency group (79% [27 out of 34]) recalled only two or less of the six target causal relations.

Table 2. *Means and standard deviations of performance on the causal question and the problem-solving test*

	Causal question		Problem solving	
Proficiency	*M*	*SD*	*M*	*SD*
High	2.75	1.38	5.47	3.26
Low	1.85	1.16	2.88	1.68

Note. Scores on the causal question and the problem-solving test range 0 to 6 and 0 to 12, respectively.

Similarly, performance on the problem-solving test was significantly better in the high- than in the low-proficiency group, $t(68) = 4.14$, $p < .001$, $d = 0.99$. It was hence confirmed that the high-proficiency group outperformed the low-proficiency one in both memory for causal relations and learning outcomes from the text.

Contributions of memory for causal relations to learning from the text

To examine the relation between memory for causal relations and text learning, I first computed correlations between causal question and problem-solving performance separately for the two proficiency groups. The results revealed quite different pictures between the groups. In the high-proficiency group, performance on the causal question and the problem-solving test were significantly correlated, $r = .68$, $p < .001$. By contrast, this correlation in the low-proficiency group was much lower and failed to reach significance, $r = .12$, $p = .490$. This intergroup difference was statistically significant, $z = 2.80$, $p = .005$.

Based on the correlation results, I next ran a hierarchical regression analysis using problem-solving performance as a dependent variable. This regression analysis was designed to clarify whether L2 reading proficiency modified contributions of causal question performance to the problem-solving. In Steps 1 and 2, L2 reading proficiency test scores (termed Proficiency) and causal question performance (Causal Memory) were entered as predictor variables, respectively. In Step 3, I entered the interaction term of proficiency and memory for causal relations (the Proficiency × Causal Memory interaction), which was created by multiplying L2 reading proficiency test scores by causal question performance. The focus was on the significance of the R^2 change associated with the entry of the Proficiency × Causal Memory interaction; this value represented the extent to which the interaction explained the variance of problem-solving performance above and beyond the main effects of proficiency and causal memory. The results did confirm the Proficiency × Causal Memory interaction, $\beta = .26$, $p = .015$, with its entry increasing the model's predictive power by 5% (Table 3).

To interpret this interaction, I performed a simple slope test. Figure 1 shows the effect of memory for causal relations on problem-solving performance as a function of readers' proficiency. It was found that better problem solving was associated with increased causal memory when readers' proficiency was high (one *SD* above the mean of L2 reading proficiency scores), $\beta = .60$, $p < .001$. On the other hand, such a link was not found when readers' proficiency was low (one *SD* below the mean), $\beta = .11$, $p = .470$. It was thus corroborated that

contributions of memory for causal relations to text learning depended on participants' L2 reading proficiency. To better understand this result, the following two sections report the results from the qualitative analyses on causal question answers and problem-solving responses.

Table 3. *Results of hierarchical regression analysis on problem-solving performance*

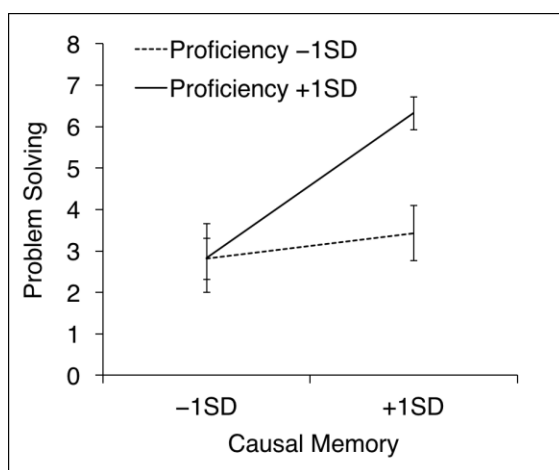| Step (Predictor)	β	R^2	ΔR^2	F for ΔR^2	p for ΔR^2
Step 1 (Proficiency)	.25*	.31	-	-	-
Step 2 (Causal Memory)	.35**	.40	.09	10.45	.002
Step 3 (Proficiency × Causal Memory)	.26**	.46	.05	6.18	.015

Note. *p < .05. **p < .01.

Figure 1. The relationship between memory for causal relations and problem-solving performance (± standard errors) as a function of L2 reading proficiency.

Inferences in causal question answers

Table 4 presents the results of the analysis on three-type inferences found in causal question answers. Due to the standard deviations being large, I used the non-parametric Mann-Whitney *U* test for statistical analyses.

Table 4. *Means and standard deviations of the number of inferences in causal question answers*

Proficiency	Total amount of inferences		Within-text inferences		Elaborative inferences		Incorrect inferences	
	M	SD	M	SD	M	SD	M	SD
High	0.75	0.73	0.56	0.61	0.08	0.28	0.11	0.32
Low	0.68	0.77	0.18	0.39	0.12	0.33	0.38	0.49

The total amount of inferences was not significantly different between the proficiency groups, *U* = 570, *z* = 0.54, *p* = .590, *r* = .06. Likewise, proportions of participants who produced at least

one inference were not significantly different between the groups: the high-proficiency group at 58% (21 out of 36) and the low-proficiency group at 53% (18 out of 34), $\chi^2(1) = 0.21$, $p = .650$, φ = 0.05. However, proficiency effects emerged within respective types of inferences. Specifically, the high-proficiency group showed more within-text inferences than the low-proficiency group, $U = 408$, $z = 2.90$, $p = .003$, $r = .35$. Oppositely, the low-proficiency group showed more incorrect inferences than the high-proficiency group, $U = 446$, $z = 2.63$, $p = .009$, $r = .31$. This trend was augmented by the fact that proportions of incorrect inferences were higher for the low- ($M = .33$, $SD = .46$) than for the high-proficiency group ($M = .08$, $SD = .25$), $U = 440$, $z = 2.70$, $p = .007$, $r = .32$. There was no significant intergroup difference for elaborative inferences, $U = 591$, $z = 0.47$, $p = .635$, $r = .06$.

Distinctive patterns of problem-solving responses in the low-proficiency group

Scrutiny into problem-solving responses revealed two distinctive patterns in the low-proficiency group. First, the low-proficiency group tended to answer events in a causally incorrect order. This was observed for 47% (16 out of 34) of the low-proficiency group and 19% (7 out of 36) of the high-proficiency group, $\chi^2(1) = 6.04$, $p = .013$, $\varphi = 0.29$. For example, a low-group participant answered the question in Table 1 that "Usually, the heart shrinks in space in order to reduce the body fluids. With that disease....". In the underlined part, the cause-effect order between the amount of body fluid and the heart size was reversed; in fact, the heart shrinkage is the *outcome* of the decreased body fluids, not the cause.

Second, the low-proficiency group tended to answer using information that is irrelevant to situations conveyed by the text. This was observed for 41% (14 out of 34) of the low-proficiency group and 11% (4 out of 36) of the high-proficiency group, $\chi^2(1) = 8.27$, $p = .004$, $\varphi = 0.34$. One low-group participant answered the question in Table 1 that "If people did not notice changes in body fluids, the amount of exercise would not differ from on the earth. So, the heart would not shrink." The relation between the amount of body fluids and the amount of exercise was not implied, let alone stated, by the text.

Discussion

Regarding RQ1 (Do contributions of L2 readers' memory for causal relations to their learning outcomes from the text differ as a function of L2 reading proficiency?), the results confirmed that contributions of memory for causal relations to text learning depended on L2 reading proficiency. Specifically, the regression results showed that causal question performance predicted the problem solving when participants' L2 reading proficiency was high but not when L2 reading proficiency was low. This is consistent with the prediction that the difficulty with inferential processing would prevent low-proficiency readers from learning causal relations from the text. The combination of quantitative and qualitative analyses in the present experiment specifically provided two explanations as follows.

The first concerns memory for causal relations, which is relevant to RQ2 (How does L2 reading proficiency affect L2 readers' memory for causal relations?). That is, low-proficiency readers had difficulty understanding causal relations in the text; the low-proficiency group correctly

answered much fewer causal relations than the high-proficiency group. It is important to note that the causal relations needed to complete the causal question were explicitly provided in the text (see Appendix A for the experimental text). Hence, answering these relations can be achieved by memory for the explicit text (i.e., textbase). From this perspective, many of the low-proficiency readers presumably had trouble at the textbase level.

The second is that low-proficiency readers failed to construct accurate situation models. This view comes from the results regarding RQ3 (How does L2 reading proficiency affect L2 readers' inferences generated in the causal question?) and RQ4 (How are low-proficiency L2 readers characterized by contents of their problem-solving responses?). Regarding RQ3, the qualitative analysis of causal question answers revealed that low-proficiency readers showed much more incorrect inferences than high-proficiency ones. A larger number of incorrect inferences observed for low-proficiency readers indicate that situation models constructed by those readers were of lower accuracy, relative to high-proficiency ones (Barry & Lazarte, 1998). Regarding RQ4, low-proficiency readers tended to perform the problem solving using inappropriate causal sequences and irrelevant information, as shown by the qualitative inspection into problem-solving responses. These patterns of the incorrect problem solving suggest that causal relations in the text were not correctly learned as long-term memory knowledge in low-proficiency readers. Consequently, those readers were inefficient at applying the causal relations beyond the text.

Combining these two accounts, it can be assumed that low-proficiency readers had trouble with processes both at textbase and situation-model levels; they struggled to understand explicit causal relations in the text and failed to construct accurate situation models. In general, the successful construction of situation models depends on the appropriate understanding of what is stated in the text (i.e., a building of the textbase). This fact implies that low-proficiency readers' difficulty was presumably rooted in processes associated with the building of the textbase. Processes at the textbase level include understanding individual words and phrases and interrelating those ideas to form a meaningful proposition (Kintsch, 1998). At this point, it must be noted that lexical and syntactic items of the experimental passage were highly simplified. Accordingly, it is unlikely that low-proficiency readers had difficulty understanding individual words or sentences per se. Instead, their difficulty seems to lie in the process of interrelating multiple ideas to form coherent propositional idea units. This argument is consistent with prior L2 research demonstrating that less-proficient readers often fail to build a relational understanding of the text, even when they can capture the meaning of text elements individually (e.g., Horiba, 2000; Hosoda, 2014). In sum, the results suggest that low-proficiency readers' difficulty with relational understanding leads to their situation models having low accuracy, which then prevented their memory for causal relations from being learned as knowledge.

In addition to the above findings regarding L2 text learning, I must note several points that require careful interpretation. First, the fact that the causal question was conducted after reading leaves open the possibility that the observed inferences were made when participants answered the question, not at the time of comprehension (i.e., during reading). Some past studies have indeed reported that when required to do so by the post-reading task, readers may make inferences that were not generated during reading (Hosoda, 2014; Noordman et al., 1992). In particular, it is possible that low-proficiency readers were forced to rely on off-text information when answering the causal question because their memory for causal relations was relatively

weak (as shown by their lower performance on this task). This implies that a larger number of incorrect inferences observed in low-proficiency readers might be an artifact of this methodological feature of the causal question. This study cannot rule out this possibility because participants' on-line reading processes were not directly measured. However, it must also be noted that this possibility does not mean that low-proficiency readers generated correct inferences during reading. Rather, considering that they struggled with even textbase processes, it was presumably difficult for low-proficiency readers to make appropriate inferences during reading. This is because the success in on-line inference generation usually requires the efficient processing of explicit text (Horiba, 1996; Kintsch, 1998).

Second, the number of inferences observed in this study was much lower than that found in previous research; even high-proficiency readers demonstrated less than one inference on average. This is, at least in part, attributable to the nature of the experimental text. The text used here was short, with highly controlled linguistic complexity. The information source for inferences was accordingly reduced, possibly restricting the number of inferences made by readers. In contrast to this study, Barry and Lazarte (1998) used longer passages[4] and elicited on average 4.9 and 8.1 inferences in recall protocols from low-knowledge and high-knowledge readers, respectively. In addition, they reported significantly more inferences ($M = 7.6$) for passages with higher syntactic complexity (defined by the number of embedded clauses per sentence) than for passages with lower syntactic complexity ($M = 5.4$). A definitive conclusion cannot be drawn because this study did not compare different versions of the text. However, it is at least possible that longer authentic texts might have elicited more inferences.

The small number of inferences in causal question answers may also be attributable to the specificity of the instruction. As opposed to recall in Barry and Lazarte's (1998) study, where participants were asked to freely reproduce text content (their instruction was to "write everything they could remember in English without looking back at the text" p. 181), the causal question specifically instructed participants to explain target causal relations, with the beginning and the end of the causal chain provided as a cue. Such a highly specific instruction might focus participants' attention on targeting causal relations that were explicitly stated in the text. As a result, these were more likely to be produced than inferences, especially in high-proficiency readers who had constructed a relatively robust memory for the key causal relations. Given the instruction's focus on explicit information, inferences observed in the causal question should be considered tentative.

Finally, the results might be affected by readers' individual differences not considered here. For example, working memory capacity might have impacted causal question performance, given that the task was administered immediately after reading. However, this study cannot provide a direct account of such potential variables, because subskills of L2 reading proficiency (e.g., prior knowledge on the text's subject matter, vocabulary knowledge, and syntactic parsing skills) were not assessed.

Presumably, the most critical factor is prior knowledge. It is widely agreed that how well readers learn from text is largely dependent on the amount of knowledge in the text's domain (e.g., Kintsch, 1994; McNamara et al., 1996). In the present study's case, prior knowledge might have interacted with L2 reading proficiency. Some researchers have shown that struggling readers

often fail to use or integrate prior knowledge with text information to deepen comprehension (e.g., McNamara & O'Reilly, 2009). Hence, high-proficiency readers were more likely to effectively use prior knowledge to learn better from the text. By contrast, simply possessing prior knowledge might have been insufficient to support low-proficiency readers' learning, especially considering that the present low-proficiency group struggled with understanding the explicit text. In general, effective knowledge use builds on the efficient operation of basic reading processes (Kintsch, 1998). It is likely that a certain level of L2 linguistic skills or knowledge is a prerequisite for making good use of prior knowledge to facilitate learning (Hosoda, 2015). An important direction for future research is to examine how L2 readers of different linguistic proficiency may or may not benefit from prior knowledge in their expository comprehension and learning.

Conclusion

The present study explored the relation between L2 readers' memory for causal relations and learning from expository text. The findings of the study indicate that memory for causal relations contributed to text learning only in readers with high L2 reading proficiency. Low-proficiency readers, on the other hand, experienced difficulty with the construction of accurate situation models, as well as the understanding of causal relations explicitly given in the text; consequently, they failed to learn causal relations in the text as knowledge applicable to a new situation.

As an implication for educators, the findings highlight the need for support for low-proficiency L2 students. It is important to note that an intervention for lexical- or syntactic-level processes alone would be insufficient to scaffold less-skilled readers to deep-level expository comprehension, as suggested by the fact that the low-proficiency group still struggled with the linguistically highly simplified passage. Rather, support for struggling L2 students should be aimed at facilitating processes involved in relational understanding. One example is the provision of pre-reading questions that guide readers' attentional allocation. According to the goal-focusing model (McCrudden, Magliano, & Schraw, 2011), readers' attentional allocation is determined by the relevance to a reading goal generated by given questions or instructions. Questions specifically targeting important relations in the text can make those relations explicit to students. For example, providing causal questions before reading would direct students' attention toward causal relations between information in text and in turn support the attainment of a causal understanding of that text. Actually, the provision of causal questions has been shown to be effective in facilitating comprehension in struggling readers (Carlson et al., 2014; McMaster et al., 2015).

Several limitations of this study should be noted. Above all, it must be emphasized that the present study employed only one passage. Future research is recommended to use several versions of authentic passages to explore the possible interplay of linguistic features of text and L2 proficiency in learning from text in L2 reading. It must also be noted that this study cannot provide direct evidence of L2 readers' inference generation during reading, as participants' inferences were assessed by an off-line task (i.e., the causal question). Direct investigation of L2 readers' online expository reading processes is needed to provide a more grounded account of L2 readers' moment-by-moment inferential processing. This can be done, for example, by using a

think-aloud method. Finally, we should consider the fact that participants read the text only once. During initial reading, less-skilled L2 readers must usually devote most of their cognitive resources to capturing the meanings of individual text elements, thus leaving few cognitive resources for relational processes (e.g., Horiba, 1996). Research exploring how L2 readers develop causal understanding through multiple readings can reveal L2 readers' potential to achieve deep-level comprehension.

Acknowledgments

Notes

1. The four problem-solving questions were prepared through a pilot study in Hosoda (2016). Six candidate questions were first constructed through discussion between the author and a graduate student majoring in English education based on three criteria: (a) correct responses require readers to make inferences from the text, but (b) do not demand technical knowledge, and (c) queried situations are not restricted to the context of the text (McNamara & Kintsch, 1996). Forty-three Japanese EFL students read the experimental text and answered the created six candidate questions. The results indicated that participants' responses to two of the six questions varied excessively. Thus, they were excluded, and the remaining four questions were retained.

2. The interdependence among causal events was not specifically considered in the scoring of the causal question. However, participants seemed to select events based on the distance to the CR1 and CR8 that were given as a cue in the instruction. Specifically, CR2 and CR7, adjacent to CR1 and CR8, were produced by the majority of the participants. Oppositely, CR3 and CR4, the most distant to CR1 and CR8, were least produced. This might have created a difference between what participants remembered from the text and what they actually provided in response to the causal question. I thank the reviewer for pointing this out.

3. The ideal answers to the problem-solving test were constructed in a four-step procedure. First, four judges (all familiar with the experimental text) independently created candidate answers to the four problem-solving questions. Second, these candidate answers were parsed into individual ideas (Japanese clauses). Third, the four judges held a discussion to identify ideas that were necessary for each answer to the problem-solving questions; ideas were deemed as necessary when three or more of the judges included in their candidate answers. Finally, the ideal answers were determined by assembling the identified necessary ideas.

4. Barry and Lazarte (1998) do not provide specific information about features of their experimental passages (e.g., the number of words or sentences). However, the first paragraph

from one of their passages is provided in Barry and Lazarte's (1995) appendix, where it can be seen that this paragraph alone is the same length as or longer than this study's passage. I thank the reviewer for this suggestion.

References

Barry, S., & Lazarte, A. A. (1995). Embedded clause effects on recall: Does high prior knowledge of content domain overcome syntactic complexity in students of Spanish? *The Modern Language Journal*, *79*, 491–504. doi: 10.1111/j.1540-4781.1995.tb05449.x

Barry, S., & Lazarte, A. A. (1998). Evidence for mental models: How do prior knowledge, syntactic complexity, and reading topic affect inference generation in a recall task for nonnative readers of Spanish? *The Modern Language Journal*, *82*, 176–193. doi: 10.1111/j.1540-4781.1998.tb01191.x

Carlson, S. E., van den Broek, P., McMaster, K., Rapp, D. N., Bohn-Gettler, C. M., Kendeou, P., & White, M. J. (2014). Effects of comprehension skill on inference generation during reading. *International Journal of Disability, Development and Education*, *61*, 258–274. doi: 1034912X.2014.934004

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

Gilliam, S., Magliano, J. P., Millis, K., Levinstein, I., & Boonthum, C. (2007). Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (R-SAT). *Behavior Research Methods*, *39*, 199–204. doi: 10.3758/BF03193148

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition*, *18*, 433–473. doi: 10.1017/S0272263100015370

Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, *29*, 223–267. doi:10.1207/S15326950dp2903_3

Hosoda, M. (2014). Generating causal bridging inferences in EFL expository reading: Combining on-line and off-line processing. *ARELE (Annual Review of English Language Education in Japan)*, *25*, 239–254.

Hosoda, M. (2015). Causal inferences during EFL reading of expository texts: Effects of two kinds of familiarity and L2 reading proficiency. *ARELE*, *26*, 205–220.

Hosoda, M. (2016). The interplay of text cohesion and L2 reading proficiency in different levels of text comprehension among EFL readers. *ARELE*, *27*, 201–216.

Ishikawa, S., Uemura, T., Kaneda, M., Shimizu, S., Sugimori, N., Tono, Y., . . . Murata, M. (2003). *JACET 8000: JACET list of 8000 basic words*. Tokyo, Japan: JACET.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*, 294–303. doi: 10.1037/0003-066X.49.4.294

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

León, J. A., & Peñalba, G. E. (2002). Understanding causality and temporal sequence in scientific discourse. In J. Otero, J. A. León & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 155–178). Mahwah, NJ: Erlbaum.

Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension

of easy and difficult texts. *Cognition and Instruction*, *18*, 525–556. doi: 10.1207/S1532690XCI1804_4

Magliano, J. P., Millis, K., the RSAT Development Team, Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning*, *6*, 131–154. doi: 10.1007/s11409-010-9064-2

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*, 377–389. doi: 10.1037/0022-0663.93.2.377

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2011). *Text relevance and learning from text*. Greenwich, CT: Information Age Publishing.

McCrudden, M. T., Schraw, G., & Lehman, S. (2009). The use of adjunct displays to facilitate comprehension of causal relationships in expository text. *Instructional Science*, *37*, 65–86. doi: 10.1007/s11251-007-9036-3

McMaster, K. L., van den Broek, P., Espin, C. A., Pinto, V., Janda, B., Lam, E., Hsu, H.-C., Jung, P.-G., Leinen, A. B., & van Boekel, M. (2015). Developing a reading comprehension intervention: Translating cognitive theory to educational practice. *Contemporary Educational Psychology*, *40*, 28–40. doi: 10.1016/j.cedpsych.2014.04.001

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43. doi: 10.1207/s1532690xci1401_1

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–288. doi: 10.1080/01638539609544975

McNamara, D. S., & O'Reilly, T. (2009). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In A. M. Columbus (Ed.), *Advances in psychology research* (pp. 113–136). New York, NY: Nova Science Publishers.

Millis, K., Magliano, J., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading*, *10*, 225–240. doi: 10.1207/s1532799xssr1003_2

Mulder, G. (2008). *Understanding causal coherence relations*. Utrecht, The Netherlands: Landelijke Onderzoeksschool Taalwetenschap (LOT).

Mulder, G., & Sanders, T. J. (2012). Causal coherence relations and levels of discourse representation. *Discourse Processes*, *49*, 501–522. doi: 10.1080/0163853X.2012.692655

Noordman, L. G. M., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *13*, 573–590. doi: 0749-596X(92)90029-W

Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). *The association between TOEFL iBT® test scores and the Common European Framework of Reference (CEFR) levels*. Princeton, NJ: Educational Testing Service.

Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*, *31*, 507–524. doi: 10.1016/0749-596X(92)90026-T

Ushiro, Y., Hamada, A., Kimura, Y., Nahatame, S., Hosoda, M., Kato, D., & Watanabe, Y. (2015). Building causally coherent mental representations and learning from expository texts among Japanese EFL readers. *JACET Journal*, *59*, 131–150.

Wittwer, J., & Ihme, N. (2014). Reading skill moderates the impact of semantic similarity and

causal specificity on the coherence of explanations. *Discourse Processes*, *51*, 143–166. doi: 10.1080/0163853X.2013.855577

Appendix A

Experimental 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 amount 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.

Appendix B

List of Key Causal Relations (CRs) in The Experimental Text

CR1: Lack of Gravity
CR2: Body fluids shift headward.
CR3: Body senses flood of fluids.
CR4: Body tries to decrease fluids.
CR5: Body eliminates more fluids and consumes less fluid
CR6: Body fluid levels are decreased.
CR7: Heart does not work normally.
CR8: Heart shrinks.

Appendix C

Problem-Solving Test (Translated From Japanese)

Instruction: You will answer questions that require applying what was written in the text to other situations. Please be sure to answer the questions as detailed as possible by maximally using what you learned from the text. Also, please provide not only conclusions or outcomes to what the questions ask, but also reasons for them in detail.

1. Imagine that the lungs and the heart sent messages in space, as stated in the text, but other parts of the body did not follow it. What would occur to the body water in this case?

2. Explain why people are advised to refrain from drinking a lot of water when their hearts are weakened.

3. Imagine that you got a disease that makes you insensitive to changes in body fluids. Explain why, in this condition, your heart size would not change even if you stayed in space.

4. How could a space station be designed so that astronauts' hearts would be less likely to shrink?

About the Author

Masaya Hosoda is a graduate student at University of Tsukuba and a research fellow at Japan Society for the Promotion of Science. His interests include L2 reading, expository text comprehension, learning from text, and understanding of causal relations. E-mail: hzm127@gmail.com