Generating Causal Bridging Inferences in EFL Expository Reading: Combining On-Line and Off-Line Processing

Masaya HOSODA
Graduate School, University of Tsukuba

Abstract

Although causal bridging inference generation has been evidenced to be instrumental in understanding and learning from expository texts, little research has been done with EFL readers. The present study therefore explored whether EFL readers generate causal bridging inferences from expository texts, focusing on two phases of inference processing: on-line and off-line. In the on-line phase, a total of 51 Japanese undergraduates read expository passages and answered verification questions. In the off-line phase, a sentence recognition task was used to evaluate off-line inference generation. The joint pattern of reading times for target sentences and response times for inference questions revealed that regardless of L2 reading proficiency, participants failed to make appropriate inferences during expository reading (i.e., on-line). However, the results of the sentence recognition task indicated that when required to engage in extended reflection after reading (i.e., off-line), participants did generate causal bridging inferences. In particular, high L2 proficiency readers showed more robust generation than low proficiency readers. Together, the on- and off-line findings emphasize the difficulty EFL learners have in making inferences from expository texts by themselves, and thereby highlight the importance of reflective tasks that elicit additional effort to make meaning from texts.

1. Introduction

Successful comprehension of an expository text allows us to learn new information from the text. Accordingly, expository texts play an essential role in both L1 and L2 educational settings. In order to learn from an expository text, we need to not only understand explicitly stated information, but also comprehend implicit meaning of the text by generating inferences. For example, consider the sentence, Beginners are frequently advised to use compact skis because they usually have difficulty in changing directions. Readers should readily identify an explicit causal relation from the causal connective because; the subordinate clause (beginners usually have difficulty in changing directions) expresses the cause of the idea conveyed by the main clause (beginners are frequently advised to use compact skis). However, a fully coherent understanding further requires readers to infer the implicit idea underlining this causality—why are compact skis...
good for those who cannot turn well? Specifically, from the explicit causal relation, readers have to infer a mediating idea, an underlying proposition that explains why the cause brings about the effect. In this example, the mediating idea is “It is easier to turn on compact skis.” This kind of inference is called a causal bridging inference because it causally bridges pieces of textual information with an implicit linking idea (Noordman, Vonk, & Kempff, 1992; Singer, Stewart, & Harkness, 1997; Singer & O’Connell, 2003).

Generating causal bridging inferences is important for expository comprehension because causality is an essential relation that describes states or phenomena in the texts (León & Peñalba, 2002). Indeed, relevant studies have empirically demonstrated that generating causal bridging inferences contributes to the construction of coherent mental models and enhances subsequent learning from the text (e.g., Vidal-Abarca, Martinez, & Gilabet, 2000).

Despite the importance of inference generation in expository reading, however, reading instructions in the English as a foreign language (EFL) context have mostly focused on understanding explicit linguistic features, rather than inferring implicit messages. One reason for this is the lack of studies examining inference generation in EFL expository reading. Motivated by these background issues, the present study aimed to fill this gap by exploring causal bridging inference generation from expository texts in EFL readers.

Of primary significance, the study took a twofold approach, examining on-line and off-line inference generation. Inferences are on-line when they are automatically generated during reading without any specific instructions (e.g., Graesser, Singer, & Trabasso, 1994). In contrast, off-line inferences are made after reading, in response to strategic tasks or relevant information that elicits extended study or reflection (e.g., Graesser et al., 2007). These two phases of inference generation accordingly reflect different comprehension processes. Hence, studying both on- and off-line processing will provide broader insights into inference processing in EFL expository reading.

1.1 On-Line Causal Bridging Inference Generation in Expository Reading

Researchers in the field of reading comprehension agree that generating inferences on-line is more difficult for expository than for other types of texts (e.g., Singer et al., 1997). This is largely because expository texts often present less familiar information, and their structures are more complex. The unfamiliarity of the text may leave readers with less of a knowledge base that is instrumental in automatic inference generation (Noordman et al., 1992). On the other hand, the structural complexity can impede readers’ search for information necessary for on-line inference processing (Wiley & Myers, 2003). Indeed, previous studies have empirically demonstrated that even L1 readers cannot easily make causal bridging inferences during expository reading.

Singer and colleagues (Singer, et al., 1997; Singer & O’Connell, 2003) used measures of reading times for target sentences and response times for verification questions to investigate on-line causal bridging inference generation with relatively familiar expository passages. They used reading materials, as shown in Table 1, containing a target because sentence, from which the
mediating idea should be inferred. Two conditions were set for each passage: an explicit and implicit. Specifically, in the explicit condition, the target sentence was preceded by an *explicit sentence* that explicitly stated the mediating idea that causally bridged the two clauses of the target sentence. The explicit sentence was removed in the implicit condition. Thus, readers in the implicit condition had to infer the implicit mediating idea to coherently understand the target sentence. After reading, readers verified an *inference question* that queried the mediating idea that was inferable from the target sentence. The logic is as follows: if readers made the inferences on-line, response times for inference questions should not differ between the implicit and explicit conditions because the questions probed the inferential mediating idea that was explicitly available in the explicit condition and should have been inferred in the implicit condition. Conversely, target reading times should be longer in the implicit condition due to the additional cognitive processes required to infer the missing idea. The results from a series of experiments showed that readers in the implicit condition did take longer on target sentences and took no longer to answer inference questions. However, this is the case only when two clauses of the target sentence were linked by the causal connective *because*. Consequently, they concluded that L1 readers can make causal bridging inferences on-line with the appropriate causal connectives signaling text causality.

In contrast, Noordman et al. (1992) demonstrated that even L1 readers cannot make the inferences during expository reading when texts are very technical. Together, these studies suggest that L1 readers of expository texts can generate causal bridging inferences on-line only when (a) text familiarity is sufficiently high, and (b) causal connectives that guide reading are available.

### 1.2 Inference Generation in L2 Reading

Although the above findings are informative, two points should be noted. First, all of these studies were conducted with L1 readers. In the context of L2, there is limited research addressing inference generation from expository texts. In particular, almost no study has focused on causal bridging inferences in EFL expository reading. However, Horiba (2000) provided some insights into differences between inference processing in expository reading in L1 and L2 readers. She used a think-aloud method where readers verbalized their thoughts while trying to make meaning from texts. The results revealed that L2 readers devote significantly more cognitive resources to lower level processing (e.g., word recognition and syntactic processing) than L1 readers, resulting in serious reduction in inference generation. Moreover, a comparison of the expository and narrative conditions suggested that this trend is even more pronounced for expository texts. Thus, these findings reveal how difficult it is for L2 readers to make inferences from expository texts. She also demonstrated that difficulty in lower level processing can impede inference processing. In this regard, Yoshida (2003) showed that because more cognitive resources are allocated to lower level processing, low proficiency L2 readers generate fewer inferences than high proficiency L2 readers. This suggests that limitations in L2 reading proficiency pose an additional challenge for inference processing in L2 readers. Together, these observations suggest that it may
be quite difficult for EFL readers to automatically generate causal bridging inferences during expository reading.

Second, prior research on causal bridging inferences has only explored on-line inference generation. Consequently, it is unclear whether readers can make causal bridging inferences off-line when on-line generation is difficult. Thus, the next question to be addressed is how do on-line and off-line inferences differ?

1.3 Off-line Inferences

Compared to on-line inferences, generating off-line inferences involves extended cognitive effort (e.g., memory retrieval) elicited by a specific task or relevant information (Graesser et al., 2007). In other words, readers are required to engage in inference processing more voluntarily or even strategically during off-line inference generation. On this account, research has demonstrated that during off-line processing, readers can generate inferences that they cannot automatically make during reading (e.g., Nahatame, 2013), and off-line inference generation is more salient than that of on-line (e.g., Inohara, Horiiuchi, & Kusumi, 2008).

Inohara et al. (2008) provided a comparison of on- and off-line inference generation in L1 narrative reading using a recognition task. First, the researchers examined on-line inference generation using an immediate recognition task that was conducted after readers processed each passage. Next, off-line inference generation was assessed using a delayed recognition that was conducted after all passages were read. The results from the two tasks indicated that readers make more inferences during off-line than on-line processing. The researchers attributed the enhanced off-line inference generation to readers' engagement in memory retrieval, which was elicited by the off-line task.

In relation to L2 readers, Nahatame (2013) explored how demands on cognitive resources affect generation of predictive inferences (inferences drawn to make a prediction about a future context) in L2 readers. He found that Japanese EFL readers cannot make predictive inferences on-line, when increased demands diminish available cognitive resources. However, even under such conditions, EFL readers were evidenced to generate the inferences off-line, when they were asked to judge whether sentences describing predictable events were reasonable in a future context. For these outcomes, he raised the possibility that the off-line task allowed readers to make the inferences in a strategic manner.

Given that off-line processing allows readers to draw inferences that are difficult to make during on-line processing, it is particularly meaningful to jointly investigate on- and off-line inference generation in the EFL context. As described in Section 1.2, on-line causal bridging inference generation from expository texts is expected to be quite challenging for EFL readers because they may experience significant difficulty in lower level processing during expository reading (Horiba, 2000). However, thanks to extended cognitive effort involved during off-line processing, it is likely that they will be able to draw difficult-to-generate inferences off-line.
Therefore, studying both on- and off-line inference processing can unveil EFL readers' potential for inference generation that might otherwise be disregarded.

2. Overview of the Present Study

The goal of the present study was to examine whether or not EFL readers generate causal bridging inferences from expository texts. In particular, the study jointly explored on- and off-line causal inference generation to attain broader insights into inference processing in EFL expository reading.

First, on-line causal bridging inference generation was investigated by measuring both reading times for target sentences and response times for inference questions (Singer et al., 1997). This paradigm was employed because it provides different and complementary perspectives on on-line inference processing (Singer et al., 1997; Singer & O’Connell, 2003). Readers were given no specific instructions to ensure that inferences were made automatically. In addition, I ensured that reading materials met the textual conditions (i.e., text familiarity and causal connective availability) reported by prior research for on-line causal bridging inference generation.

Second, off-line inference generation was assessed with a sentence recognition task (Muramoto, 2000; see Section 3.2, 3.3, & 3.4 for details). The task was unexpectedly conducted after all the passages were processed in order to engage readers in extra reflection on the texts they had read. The following two research questions (RQs) were addressed:

RQ1: Can EFL readers generate causal bridging inferences from expository texts on-line?
RQ2: Can EFL readers generate causal bridging inferences from expository texts off-line?

3. Method

3.1 Participants

Participants were 51 Japanese undergraduates majoring in agriculture, engineering, international studies, and literature. Following Singer et al. (1997), data from four participants were removed because their error rates for verification questions exceeded 33%. Data from the remaining 47 participants were analyzed. They were divided into two proficiency groups based on their L2 reading proficiency test scores. Twenty-four participants formed the high group. They scored significantly better ($M = 14.96$, $SD = 3.16$, $Min = 12$, $Max = 23$) than the other 23 participants, who formed the low group ($M = 8.09$, $SD = 2.91$, $Min = 3$, $Max = 11$), $t(45) = -7.76, p < .001, r = .76$.

3.2 Materials

a. L2 reading proficiency test: To measure participants' L2 reading proficiency, a reading test with 26 items was created. Passages were derived from the pre-first and second grades of the STEP test
(Obunsha, 2005a, 2005b). The reliability of the test was acceptable on excluding two items, Cronbach’s α = .81.

b. Reading materials: First, a total of 27 expository passages from Singer et al. (1997) were examined in a pilot study to confirm that mediating ideas were inferable from each of the target sentences: Thirteen Japanese graduate students majoring in English education rated the extent to which each mediating idea (e.g., *It is easier to turn on the shorter compact skis*) could be logically inferred from the corresponding target sentence (e.g., *Beginners are frequently advised to use compact skis because they usually have difficulty in changing directions*) on a 7-point scale (1 = *totally unnatural, 7 = very natural*). As a result, two passages with mean ratings less than 5 were eliminated, $M = 5.83, SD = 0.52$.

Next, following the view that the unfamiliarity of texts impairs on-line inference generation (Noordman et al., 1992), two Japanese graduate students majoring in English education and I had a discussion to expose overly unfamiliar passages. This led to the elimination of five passages that were regarded as too technical for EFL readers. Indeed, participants evaluated the familiarity of each mediating idea on a 7-point scale, and the mean familiarity rating was $M = 4.75, SD = 1.79$. Therefore, the familiarity of the passages used in the experiment was deemed intermediate.

Finally, low-frequency words (level 5 or over) according to *JACET 8000* (JACET, 2003) were paraphrased using higher-frequency words. Several technical terms that could not be rephrased using simpler words were explained in Japanese in parentheses (e.g., *Ultrasonic waves* (超音波); see Appendix for an example). However, no target sentences in the experimental passages contained such words. Consequently, 20 passages were selected for the experiment. Ten passages were designated as experimental passages, and the others were as control passages.

An example experimental passage was presented in Table 1. Each passage contained a target sentence, in which *because* signaled a causal relation. Following Singer et al. (1997), there was an implicit or explicit condition for each passage. In the explicit condition, the target sentence was preceded by an explicit sentence that directly referred to the mediating idea that causally bridged the two clauses of the target sentence. The explicit sentence was excluded in the implicit condition. Readers in the implicit condition therefore needed to infer the mediating idea by themselves to achieve a coherent understanding of the passage.

Each experimental passage was followed by two verification questions: an inference and a detail. The inference question probed the mediating idea that was inferable from the target sentence. The detail question queried a specific detail of the passage to encourage readers to read each sentence carefully. The correct answer was “yes” for half of the detail questions, and “no” for the other half. For the control passages, the inference question was replaced with another detail question, for which the correct answer was always “no” to prevent readers from thoughtlessly answering “yes” to the first questions. These questions were written in Japanese so that L2 decoding ability would not affect response times.
Two material sets containing five experimental and five control passages were created. The assignment of passages to conditions was counterbalanced between the sets.

<table>
<thead>
<tr>
<th>Table 1 An Example Experimental Passage and Verification Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The presentation order in the explicit condition: (a)→(b)→(c)→(d)→(e)→(f)</td>
</tr>
<tr>
<td>• The presentation order in the implicit condition: (a)→(c)→(d)→(e)→(f)</td>
</tr>
<tr>
<td>(a) Skiers of different abilities need different equipment from compact to large.</td>
</tr>
<tr>
<td>(b) It is easier to turn on the shorter compact skis. (explicit)</td>
</tr>
<tr>
<td>(c) Beginners are frequently advised to use compact skis because they usually have difficulty in changing directions. (target)</td>
</tr>
<tr>
<td>(d) Once they can control their movement, they can quickly advance in skill.</td>
</tr>
<tr>
<td>(e) コンパクトなスキーよりかさばりやすい (inference question)</td>
</tr>
<tr>
<td>(f) スキーの初心者は上達に時間がかかる (detail question)</td>
</tr>
</tbody>
</table>

Note. Reading and response times for the underlined sentence and question were recorded.

c. Recognition sentences: Following Muramoto (2000), three types of recognition sentences were written in Japanese for the five implicit-control passages: (a) Verbatim sentences that described an idea that was explicitly stated in the passages, (b) inference sentences that expressed a mediating idea that readers were expected to infer from the target sentence, and (c) incorrect sentences that described an idea neither mentioned in nor inferable from the texts (see Appendix for an example).

The reason for selecting implicit-control passages was that the other passages contained inference information (i.e., mediating ideas) in the form of explicit sentences (in the explicit condition) or inference questions (in the experimental passages). In constructing verbatim sentences, target sentences were avoided in order to ensure the independence of inference sentences. Incorrect sentences dealt with the same topic as the passage, but described it from a different viewpoint. To confirm the classification of each sentence, three Japanese graduate students majoring in English education judged types of each recognition sentence. This yielded very high reliability, Cronbach’s α = .95.

After reading all passages, participants were given these sentences and asked to answer “yes” if they thought a sentence had been stated in the texts, and “no” otherwise. A primary focus is on inference sentences. If readers make inferences off-line, they should falsely recognize inference sentences and provide “yes” responses because the corresponding inference information should be activated in their memory.

Recognition sentences were accompanied by confidence ratings that were made on a scale from 1 (totally unconfident) to 4 (very confident). The confidence ratings identified the confidence with which readers made their response (i.e., “yes” or “no”) for each recognition sentence (see Section 3.4 for how off-line inference generation was assessed).

d. Familiarity ratings: To assess the familiarity of the expository passages, mediating ideas for all the target sentences were re-written in Japanese by the author. They were coupled with 7-point familiarity ratings that ranged from 1 (totally unfamiliar) to 7 (very familiar).
3.3 Procedure

Participants were tested individually in a single session that lasted approximately an hour. The session consisted of two main phases: an on-line and off-line. In the on-line phase, on-line inference processing was assessed using passage reading and verification questions. In the off-line phase, the off-line generation of causal bridging inferences was measured with a sentence recognition task. Participants took the L2 reading proficiency test before the on-line phase, and familiarity ratings were collected later.

(1) On-line phase (passage reading and verification questions)

After the 30-minute L2 reading proficiency test, the on-line phase began. In this phase, participants read the ten experimental and ten control expository passages sentence by sentence on a computer screen and answered verification questions. Passages were presented in a random order. After the instructions were given and participants practiced reading, the signal “ready?” was presented at the center of the screen. Pressing the “yes” button on the Response Pad RB-730 initiated the first sentence. Participants were asked to read each sentence carefully and press the “yes” button to signal that they understood the sentence displayed. Reading time for the target sentence was recorded with Super Lab 4.5. After a passage was finished, the screen went blank for 2,500 ms, followed by a fixation (*** for 500 ms. Then, an inference (or a detail for control passages) question was displayed. Participants were instructed to verify whether the question was true or not as quickly and accurately as possible by pressing “yes” or “no.” The response and response time for inference questions were recorded. Following another blank interval (2,500 ms) and fixation (500 ms), a detail question appeared. Answering this led to a 1000-ms blank that was followed by another “ready?” prompt. This sequence was repeated for the 20 passages.

(2) Off-line phase (sentence recognition task)

Following the on-line phase, an unexpected sentence recognition task was conducted. Participants were provided with a booklet containing five sets of the three types of recognition sentences with confidence rating scales, and 20 mediating ideas for all the target sentences with familiarity rating scales. Participants were instructed to select “yes” if they thought each sentence had appeared in the passages, and “no” otherwise. They were also asked to select “no” for sentences that they thought could be suggested by the passage context, but were not explicitly stated. In addition, participants evaluated their degree of confidence (i.e., confidence ratings) for each recognition response using a 4-point scale. Following the recognition task, they evaluated the familiarity of the 20 mediating ideas for all the target sentences on a 7-point scale. They were asked to provide ratings based on whether they had known each idea before reading the passages.

3.4 Scoring and Data Analysis

For the on-line phase, the joint pattern of response times for inference questions and reading
times for target sentences was used to identify on-line causal bridging inference generation (Singer et al., 1997): If readers generate the inferences on-line, inference response times should be almost the same across the implicit and explicit conditions because the question queried the inference information. By contrast, target reading times should be longer in the implicit condition because of the mental effort needed to infer the missing idea.

Analyses of variance (ANOVAs) were performed on (a) reading times for target sentences, (b) correct response times for inference questions, and (c) error rates for inference questions (to evaluate inference accuracy). The between-participant variable was L2 reading proficiency (high, low), and the within-participant variable was the condition (explicit, implicit). Target reading times were divided by the number of syllables in the corresponding target sentences. Inference response times were converted into per syllable in the same way.

For the sentence recognition task, a recognition rating ranging from 0 to 6 was produced for each recognition sentence, based on the response (i.e., “yes” or “no”) and confidence rating (i.e., from 0 to 4). When the confidence rating was 1 (totally unconfident), the recognition rating became 3 regardless of whether the response was “yes” or “no.” Based on this, for “yes” responses, the recognition rating increased as the confidence rating increased. In contrast, for “no” responses, the recognition rating decreased as the confidence rating increased. For example, if a reader made a “yes” response to a recognition sentence with a confidence rating of 2, the resulting recognition rating was 4 (3 plus 1). Conversely, if the reader made a “no” response to another sentence with the same confidence rating (i.e., 2), the recognition rating was 2 (3 minus 1).

An ANOVA was conducted on recognition ratings. The between-participant variable was L2 reading proficiency (high, low), and the within-participant variable was sentence (verbatim, inference, incorrect). In line with previous recognition studies (e.g., Inohara et al., 2008; Muramoto, 2000), significantly higher recognition ratings for inference than incorrect sentences served as an indicator of off-line inference generation.

A significance criterion of $\alpha = .05$ was used throughout. Marginally significant effects were regarded as not significant.

4. Results and Discussion

4.1 On-Line Phase Results
(1) Response times and error rates

Mean correct response times and error rates for inference questions are displayed in Table 2. For inference response times, a $2 \times 2$ mixed ANOVA revealed a significant main effect of condition, $F(1, 45) = 5.23, p = .027, \eta_p^2 = .10$. Neither the main effect of proficiency nor the interaction reached significance, $F(1, 45) = 0.32, p = .576, \eta_p^2 = .007$, and $F(1, 45) = 0.14, p = .714, \eta_p^2 = .003$, respectively. This indicates that regardless of L2 reading proficiency, participants took more time to verify inference questions in the implicit than explicit condition.
(Figure 1). The analysis of inference error rates also revealed a main effect of condition, $F(1, 45) = 13.32, p = .001$, $\eta^2_p = .23$. Neither the main effect of proficiency nor the interaction was significant, $F(1, 45) = 3.20, p = .080$, $\eta^2_p = .07$, $F(1, 45) = 1.39, p = .244$, $\eta^2_p = .03$, respectively. Hence, both high and low proficiency readers made more errors for inference questions when the mediating ideas were missing (i.e., implicit condition) than when they were explicitly available (i.e., explicit condition).

(2) Reading times

Table 3 reports mean reading times for target sentences. The $2 \times 2$ mixed AVOVA yielded significant main effects of condition, $F(1, 45) = 18.63, p < .001$, $\eta^2_p = .29$, and proficiency, $F(1, 45) = 11.58, p = .001$, $\eta^2_p = .21$. More importantly, the condition $\times$ proficiency interaction was significant, $F(1, 45) = 5.42, p = .025$, $\eta^2_p = .11$. Post hoc comparisons were used to examine differences in reading times between the two groups, revealing that the high group took significantly longer to read target sentences in the implicit than explicit condition, $p < .001$. Conversely, in the low group, there was no significant difference in target reading times between conditions, $p = .191$ (Figure 2).

Table 2 Mean Correct Response Times for Inference Questions (in milliseconds)

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Explicit</th>
<th>Implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>High (n = 24)</td>
<td>136.53 (.033)</td>
<td>147.67 (.10)</td>
</tr>
<tr>
<td>Low (n = 23)</td>
<td>142.37 (.044)</td>
<td>157.78 (.14)</td>
</tr>
<tr>
<td>Total (N = 47)</td>
<td>139.39 (.038)</td>
<td>152.62 (.12)</td>
</tr>
</tbody>
</table>

Note. Error rates are in parenthesis.

Table 3 Mean Reading Times for Target Sentences (in milliseconds)

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Explicit</th>
<th>Implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>High (n = 24)</td>
<td>310.55</td>
<td>415.11</td>
</tr>
<tr>
<td>Low (n = 23)</td>
<td>430.41</td>
<td>461.72</td>
</tr>
<tr>
<td>Total (N = 47)</td>
<td>369.20</td>
<td>437.92</td>
</tr>
</tbody>
</table>

Figure 1. Response times (ms) for inference questions.

Figure 2. Reading times (ms) for target sentences.
4.2 Discussion of the On-Line Phase Results

Researchers have reasoned that if causal bridging inferences are made on-line, inference response times should not differ between the implicit and explicit conditions. In contrast, target reading times should be longer in the implicit than explicit condition. Based on this logic, this sub-section provides separate discussions of on-line causal bridging inference generation in low and high proficiency readers.

(1) Low proficiency readers

For low proficiency readers, the opposite pattern was obtained: They took significantly longer to answer inference questions in the implicit than explicit condition, whereas no significant difference was found in target reading times between conditions. Moreover, inference error rates were higher in the implicit condition, meaning that they failed to draw appropriate mediating ideas during reading target sentences. Taken together, these results demonstrate that it is too difficult for low proficiency readers to automatically generate causal bridging inferences during expository reading. Their failure to make on-line inferences might be attributable to surface-level comprehension difficulty: There was no significant difference in error rates for detail ($M = .13, SD = .07$) and implicit-inference questions ($M = .17, SD = .19$), $t(22) = 1.15, p = .264, r = .24$. This indicates that low proficiency readers experienced difficulty in comprehending explicit details and implicit ideas to the similar extent. Presumably, this surface-level comprehension deficiency restricted inference processing during reading. This is consistent with the view that L2 readers experience greater difficulty in lower level processing for expository than narrative texts (Horiba, 2000). In addition, the fact that low proficiency readers took almost the same amount of time to read target sentences irrespective of the availability of mediating ideas (i.e., across conditions) implies that they failed to apply those ideas to understanding subsequent target sentences. In other words, they failed to identify a relationship between adjacent sentences. This suggests that less proficient readers can fall into sentence-by-sentence reading, even with the short and simplified expository passages used in this study. Collectively, these findings highlight the difficulty less proficient readers experience in inferring implicit meaning during expository reading.

(2) High proficiency readers

In high proficiency readers, a more complex picture of on-line inference processing emerged. For inference response times, results were the same as for the low proficiency readers: Response times were longer in the implicit condition, suggesting that on-line causal bridging inferences were not generated. In addition, higher error rates in the implicit condition indicate that it was difficult even for proficient readers to draw the pertinent mediating ideas during reading. In contrast, target reading times for high proficiency readers interestingly show a different trend. That is, they took significantly longer to read target sentences in the implicit condition, consistent with on-line inference generation. Although prior studies have never observed this pattern of results, there are some possible explanations, described below.
One possibility is that high proficiency readers generated vague or incorrect inferences during reading. Specifically, when an implicit-target sentence was encountered, they may have detected a causal gap, and thus, engaged in cognitive effort to infer the missing information necessary to fill the gap. Increased implicit-target reading times might reflect this process. However, they could not specify the inferred information to the same extent as the inference question, or arrived at an inappropriate inference. Consequently, they had to take longer to answer implicit-inference questions and had corresponding higher error rates.

Another possibility is that inference response times and target reading times differed in their sensitivity to on-line inference processing. In other words, indexing on-line inference generation as no difference in inference response times between conditions might be a strict standard. This is because response times should not be identical across conditions, unless the exact same mediating idea was drawn. In contrast, evaluating inference generation by differences in target reading times between conditions might be a more lenient criterion as implicit-target reading times may increase when readers detect incompatibility in implicit-target sentences without drawing inferences. Following this explanation, on-line processing for high proficiency readers may have included inconsistency detection, but not inference generation.

However, based on the results from the on-line phase alone, both of these explanations are only speculative. Hence, we need to consider the off-line results to develop a more comprehensive picture of inference processing in EFL expository reading.

### 4.3 Off-Line Phase Results

Table 4 displays mean recognition ratings for each type of recognition sentence. A $2 \times 3$ mixed ANOVA revealed significant main effects of sentence and proficiency, $F(2, 90) = 208.41, p < .001$, $\eta^2 = .82$, and $F(1, 45) = 7.46, p = .009$, $\eta^2 = .14$, respectively. However, these effects were mitigated by the significant proficiency $\times$ sentence interaction, $F(2, 90) = 3.17, p = .047$, $\eta^2 = .07$. Bonferroni-adjusted post hoc analyses were performed to interpret this interaction. The critical results can be summarized as follows: (a) In both proficiency groups, recognition ratings were significantly higher for inference than incorrect sentences, all $ps < .001$ (Figure 3); (b) inference recognition ratings were significantly higher for the high group ($M = 4.12, SD = .95$) than the low group ($M = 3.24, SD = .86$), $p = .002$; and (c) for the high group, recognition ratings did not significantly differ between inference ($M = 4.12, SD = .95$) and verbatim sentences ($M = 4.72, SD = .90$), $p = .072$. 

![Figure 3. Recognition ratings for recognition sentences.](image)
Table 4 Mean Recognition Ratings for Recognition Sentences

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Verbatim</th>
<th>Mean Recognition Ratings</th>
<th>Inference</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (n = 24)</td>
<td>4.72</td>
<td>.90</td>
<td>4.12</td>
<td>.95</td>
</tr>
<tr>
<td>Low (n = 23)</td>
<td>4.62</td>
<td>.87</td>
<td>3.24</td>
<td>.86</td>
</tr>
<tr>
<td>Total (N = 47)</td>
<td>4.67</td>
<td>.88</td>
<td>3.68</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note. Recognition ratings range from 0 to 6.

4.4 Combined Discussion of Off- and On-Line Results

Overall, higher recognition ratings for inference than incorrect sentences for both proficiency groups indicate that participants generated causal bridging inferences off-line. Given the findings from the on-line phase, it is proposed that whereas on-line causal bridging inference generation from expository texts is too demanding for EFL readers, they can make the inferences when an off-line task requires them to do so. This is consistent with the observation that inference generation becomes more robust during off-line processing (Inohara et al., 2008).

In particular, high proficiency readers showed more robust generation than low proficiency readers, as reflected by higher inference ratings. This fact is not only compatible with the proficiency effects on inference generation reported in prior research (Yoshida, 2003; Muramoto, 2000), but also supports the possibility that proficient readers made vague inferences during expository reading (see Section 4.2). That is, their high inference ratings may be due to maintenance of inference information drawn during on-line processing. More specifically, they may have made some, albeit incomplete, inferences during reading (as reflected by increased implicit-target reading times), and evaluating the occurrence of inference sentences in the text after reading activated corresponding inference memory, leading to high confidence "yes" responses. Therefore, the combined findings for high proficiency readers suggest it is important to give a specific off-line task that elicits extended reflection on the text to facilitate off-line inference generation, thereby making the products of on-line cognitive effort more durable.

In contrast, low proficiency readers had to generate off-line causal bridging inferences from scratch because, as suggested by the results of the on-line phase, they seemed to generate no inferences during on-line processing. This is reflected by lower inference recognition ratings for low compared to high proficiency readers. Although less salient, a paired t test comparing "yes" response rates (calculated by dividing the number of "yes" responses by the total number of responses) for inference and incorrect recognition sentences revealed significant difference between the two sentences, t(22) = 9.44, p < .001, r = .90. This indicates that low proficiency readers falsely recognized inference sentences more frequently (M = .55, SD = .19) than incorrect ones (M = .06, SD = .13), providing supplementary evidence for off-line inference generation. Hence, the on- and off-line results suggest that even less proficient readers who experience significant difficulty in lower level comprehension of expository texts can generate
difficult-to-make inferences off-line with support from a reflective task. Further, given that they may have difficulty in even identifying a relation between adjacent sentences during reading (see Section 4.3), engaging in extended reflection after reading may be a particularly effective approach to deepen their expository comprehension.

5. Conclusion

The present study examined on- and off-line causal bridging inference generation in EFL expository reading. First, the joint pattern of target reading and inference response times revealed that it is quite difficult for EFL readers to make causal bridging inferences during expository reading (i.e., on-line; RQ1). However, the difficulty varies depending on L2 reading proficiency: Whereas no evidence was obtained for on-line inference generation in low proficiency readers, high proficiency readers possibly generated vague inferences during reading. Second, recognition ratings from the sentence recognition task indicated that EFL readers can generate causal bridging inferences off-line by engaging in extended reflection after reading (RQ2). In addition, high proficiency readers generated off-line inferences more robustly, revealing the same proficiency effects as prior research (e.g., Yoshida, 2003). Combining the on- and off-line findings, we can conclude that although it is challenging for EFL readers to generate pertinent causal bridging inferences during expository reading by themselves, they do have the potential for making these inferences when provided with an off-line task that elicits extra cognitive effort.

Furthermore, the results of the present study may have some pedagogical implications. First, as participants failed to infer appropriate mediating ideas during expository reading, it is presumably difficult for EFL readers to understand expository texts beyond explicit textual information. This emphasizes the importance of teacher intervention in EFL expository reading: It would be helpful to give specific reading goals, activate students’ prior knowledge, or improve text cohesion when having students make meaning from expository texts. However, these approaches may still be insufficient for poor learners to infer implicit messages from the text: They may have difficulty in linking even adjacent sentences during expository comprehension, as suggested by the fact that their target reading times were not faster when mediating ideas were available. For such students, the findings from the off-line phase suggest a useful approach: By conducting a reflective task after reading, teachers can strategically lead even less proficient students to derive what the texts implicitly communicate. For example, having students explain text content after reading (e.g., Explain why compact skis are good for beginners) can facilitate additional thinking about the text, and thereby allow for deeper understanding of expository texts.

Finally, we must acknowledge that the present study has several limitations that should be addressed in future research. The most notable is that the passages used in the experiment were very short: They consisted of only four sentences in the explicit condition, and three sentences in the implicit condition. However, in reality, EFL learners face much longer expository texts
containing several hundred words or more. Inference generation from longer texts may additionally be affected by readers' working memory capacity, or text structure (e.g., Whitney, Ritche, & Clark, 1991). Therefore, further research should clarify the effects of reader and textual factors on inference processing involved in understanding longer expository texts.

Another limitation is that this study only used quantitative approaches. Future research using qualitative methods (e.g., a think-aloud or an eye-tracking measure) should identify the specificity and time-course of inferences. We can further expect such studies to specify differences in reading strategies between those who successfully make inferences and those who do not. Addressing these issues will provide more practical insights into EFL expository comprehension and contribute to developing EFL students' ability to learn from texts.

Acknowledgements

I would like to thank Prof. Yuji USHIRO and my colleagues for their valuable comments. I also extend my gratitude to Dr. Murray Singer for providing experimental materials. Finally, I wish to thank three anonymous reviewers for their insightful comments to improve this paper.

References


Muramoto, T. (2000). Daini gengo no bunsho rikai katei ni oyobosu shuujikudo no eikyo [The
effects of second-language proficiency on text comprehension]. *Science of Reading*, 44, 43–50.


**Appendix: An Implicit-Control Passage, Detail Questions and a Set of Recognition Sentences**

(a) Ultrasonic waves (超音波) are high frequency sound waves, having frequencies above 20,000 Hz.

(b) Ultrasonic waves are called silent sound because they have such high frequencies (target).

(c) Researchers in the field originally used another term, ‘supersonics’, for these waves.

(d) 超音波の周波数は約 10,000Hz だ (detail question)

(e) 超音波にはかって異なる名称が用いられていた (detail question)

(1) Verbatim Sentence: 超音波は周波数の高い音波だ
(2) Inference Sentence: 周波数の高い音は人の耳には聞こえない
(3) Incorrect Sentence: 超音波を利用して獲物を捕まえる動物もいる