Suppressing and Revising Incorrect Predictive Inferences in EFL Reading: An Empirical Study Using the Meaningfulness Judgment Task and Sentence Recognition Task

Shingo NAHATAME
Graduate School, University of Tsukuba / The Japan Society for the Promotion of Science

Abstract

Studies have shown that L2 learners make predictions about what might be written in the subsequent context (i.e., activate predictive inferences) during reading. However, learners do not always activate appropriate predictive inferences that are consistent with the subsequent context. Thus, the present study was conducted to examine how EFL readers deal with initially activated inferences when they are disconfirmed by the subsequent context, taking into consideration the methodological issues of previous studies. In the experiment, 43 Japanese university students read several short narratives designed to elicit an inference about the predicted event or to disconfirm the activated inference. Immediately after reading each passage, participants performed the meaningfulness judgment task (MJT) to a target sentence describing the predicted event. They also engaged in the sentence recognition task (SRT) to target sentences after reading all the passages. The response times of the MJT demonstrated that the activation of predictive inferences was not suppressed immediately after reading the context disconfirming the inferences. In contrast, the results of the SRT revealed that readers partially revised the disconfirmed inferences in their text memory. These results suggested that EFL teachers should pay attention to not only students’ predictions in text reading, but also the suppression and revision of their predictions.

1. Background

In foreign language reading instruction, teachers often ask learners what might happen next in the situation described in a text. In fact, researchers suggest that asking this question helps learners to obtain a deeper understanding of the text and become interactive readers (Day & Park, 2005; Tanaka, Shimada, & Kondo, 2011). In addition, learners sometimes spontaneously make inferences about what might happen next during reading (Horiba, 1996; Yoshida, 2003); these are called predictive inferences. For example, when reading the sentence No longer able to control his anger, the husband threw the delicate porcelain vase against the wall (Klin, Guzmán, & Levine,
1999), most readers would predict "the vase was broken" as a highly probable outcome of the described event; they activate a predictive inference.

However, readers do not always make accurate predictions. Sometimes, when they read the subsequent text, they learn that their predictions were incorrect. Consider that the aforementioned example text is followed by a sentence disconfirming the prediction, such as By sheer luck, the vase hit at an angle such that it was not damaged (Potts, Keenan, & Golding, 1988). In this case, the readers' initial prediction of the vase breaking is not supported; therefore, it is necessary to modify the prediction to achieve appropriate comprehension of the text. According to Ushiro (2010), active and flexible modification of comprehension is key for learners to be successful readers. Therefore, this study aimed to investigate such flexible text comprehension processes of Japanese learners of English as a foreign language (EFL), with a special focus on learners' spontaneous predictions (i.e., predictive inferences) during reading. The following sections review previous studies on predictive inferences in reading.

1.1 Activation of Predictive Inferences in Reading

Of several types of inferences produced in text reading, such as inferences about causal antecedents of described events or referential words, predictive inference has received considerable attention from numerous researchers. Predictive inference is the anticipation of the likely outcome of an event described in a narrative text. Predictive inferences benefit readers in ways such as easing the processing of the subsequent context, promoting the construction of a situation model, and encouraging active engagement with the text (Linderholm, 2002).

Previous studies revealed that these inferences are automatically activated during first language (L1) reading. Several factors affect the activation: the degree to which the text constrains a possible inference (e.g., Klin, Guzmán et al., 1999), whether inferences are related to motivation for a character's action (e.g., Klin, Murray, Levine, & Guzmán, 1999), and the reading purpose or strategy (e.g., Allbritton, 2004). Furthermore, some researchers showed that activated predictive inferences are encoded as part of readers' text memory (e.g., Klin, Muuray et al., 1999).

Studies on second language (L2) reading found that L2 learners also activate predictive inferences during reading. However, some studies demonstrated that predictive inferences are more likely to be drawn by learners with high L2 proficiency than those with low L2 proficiency (Horiba, 1996; Yoshida, 2003). Moreover, Nahatame (2012) suggested that predictive inferences are more likely activated during L2 reading when the inferences are strongly constrained by the context and related to the motivation for a character's action (see also 2.2.1 Passages). These past studies on L1 and L2 reading have investigated inference activation by analyzing participants' response times to target words or sentences representing the inference concept (e.g., break; The vase was broken) or think-aloud protocols during the reading of experimental passages.
1.2 Suppression of Predictive Inferences During Reading

As reviewed above, several studies explored predictive inferences in L1 and L2 reading and revealed that these inferences are more likely to be activated under certain conditions. In contrast, few studies have examined how readers deal with initially activated inferences when the subsequent context disconfirms the inferences. When the activated inferences are disconfirmed, readers must suppress the activation and then revise these inferences (Gernsbacher, 1997).

Suppression during reading is defined as decreased activation of text information that is inappropriate or less relevant to current comprehension. Potts et al. (1988) examined the suppression of predictive inferences of L1 English readers. They compared the strength of target inference activation while reading (a) predictive texts, in which target predictive inferences (e.g., break) are suggested (e.g., No longer...), and (b) disconfirming texts, in which suggested inferences are disconfirmed by the subsequent context (e.g., No longer...By sheer...). Lexical decision and naming times to target words did not significantly differ between the two text types, suggesting that inference activation was not suppressed even after processing the context disconfirming the inferences. Similarly, in Iseki (2006), L1 Japanese readers read predictive and disconfirming texts and then performed the meaningfulness judgment task (MJT), in which they decided whether target sentences describing inferred events made sense or were semantically acceptable. The analysis of judgment times yielded a result opposite to that of Potts et al. (1988); inference activation was suppressed for the disconfirming texts. Further, Nahatame (2012) conducted a similar experiment with Japanese EFL learners using a word verification task, in which participants were required to verify whether the target word had appeared in the text. The analysis of verification times suggested that the pattern of inference activation was not different between the predictive and disconfirming texts, consistent with the findings of Potts et al. (1988).

| Table 1 Summary of Previous Findings Regarding the Suppression of Predictive Inferences |
|-----------------------------------------------|-----------------------------------------------|-----------------|
| Task                                         | Target                                         | Inference activation |
| Potts et al. (1988)                          | L1                                             | Predictive = Disconfirming |
| Iseki (2006)                                 | L1                                             | Predictive > Disconfirming |
| Nahatame (2012)                              | EFL                                            | Predictive = Disconfirming |

As summarized in Table 1, previous studies of the suppression of predictive inferences have yielded mixed results. One possible reason behind the different patterns of results is variations in the tasks used to access inference activation; task characteristics might have affected the results of past studies. Importantly, Iseki (2006), who showed the suppression of inferences, suggested that the task can be more sensitive to the degree of inference activation when the presented target is a sentence than when it is a word, due to a greater degree of overlap between a target and an inferential concept. Thus, the MJT is assumed to be a better index of the suppression of inferences than other tasks such as lexical decision or word verification tasks. Given such a difference between tasks, the suppression of predictive inferences during EFL reading seems to require...
further examination by using a task different than that used in Nahatame (2012). Thus, the present study adopted the MJT and aimed to reexamine the results of Nahatame.

Porte (2012) described the importance of replication studies in L2 research as a means to confirm and better understand the findings of the original studies. Similarly, Jiang (2012) emphasized the importance of follow-up experiments using different tasks; such experiments “help determine if a finding is an outcome of adopting a specific task or reflects a more general phenomenon” (p. 77). Therefore, if the MJT is used and the results are different than those of Nahatame (2012), it is probable that the assessment of suppression is affected by the task used. On the other hand, if the MJT is used and the results are consistent with those of Nahatame, converging evidence from the different tasks strengthens the validity of the previous findings.

1.3 Revision of Predictive Inferences After Reading

Revision of predictive inferences was also investigated in this study, apart from suppression during reading. According to Rapp and Kendeou (2007), revision is a specific form of text comprehension updating that requires readers to replace incorrect or old text information with the correct or new information. The present study examined EFL readers’ revision of disconfirmed predictive inferences in text memory (i.e., complete text representations after reading). If incorrect inferences are encoded in memory without revision, it can lead to comprehension failures. Thus, it is important to investigate whether and how inferences are revised after they are disconfirmed.

Despite its importance, however, few studies have explored the revision of predictive inferences. Nahatame (2012) attempted to examine EFL readers’ text memory after reading predictive and disconfirming texts by comparing how much predictive information was produced in their written recall protocols. The predictive and disconfirming texts resulted in the production of the same amount of predictive information, suggesting that disconfirmed inferences remained as part of text memory. However, one possible problem with the recall analysis is the difficulty conclusively determining whether readers revised the disconfirmed inferences. When participants read the disconfirming texts, some of them recalled inferences as if they falsely believed that the inferred events actually happened in the described situation (e.g., *He threw the vase and broke it*), while others referred to their inferences just to explain the characters’ action (e.g., *He threw the vase to break it*). In the latter case, it was unclear whether the participants believed that the inferred event was what actually occurred; that is, whether they revised the inferences.

To address this issue, this study attempted to directly investigate the revision of inferences by using a sentence recognition task (SRT) (Muramoto, 2000). In this task, after reading several experimental passages, participants are presented with target sentences describing inferred events (e.g., *The vase was broken*) and then asked to judge whether the sentences are written in the passages. If participants read the disconfirmed texts and then successfully revise the target inferences in their memory, the target sentence should be recognized as not written in the text. A more detailed explanation of this task is provided in the Methods section.
1.4 Purpose and Hypotheses of the Present Study

As reviewed above, only a few studies have investigated the suppression and revision of predictive inferences in reading; unfortunately, their findings were contradictory or unclear. Therefore, the purpose of the present study was to provide further evidence regarding the suppression and revision of the activated but incorrect predictive inferences in EFL reading, considering the characteristics of the experimental tasks. The following two hypotheses (Hs) from Nahatame (2012) were examined:

H1 The activation of predictive inferences is not suppressed during EFL reading even when the following context disconfirms the inferences.
H2 Predictive inferences are not revised in EFL readers’ text memory even after they read the context disconfirming the inferences.

2. Methods

2.1 Participants

Participants were 43 Japanese undergraduate and graduate students with a variety of majors (26 female and 17 male; aged 18–25 years, $M = 20.52, SD = 1.80$). All participants had studied English as a foreign language for more than six years, and they were assumed to have approximately intermediate-level English proficiency.

The data from seven participants who did not follow the instructions or had an error rate that exceeded 25% on an experimental task (comprehension questions or the MJT) were omitted. Therefore, the data analyses were based on 36 participants.

2.2 Materials

2.2.1 Passages

Sixteen short narrative passages were adopted from Virtue, van den Broek, and Linderholm (2006) and modified slightly. Most of them were also used in Nahatame (2012). As these passages had been designed for native English-speaking readers, the sentence structure was simplified and low-frequency words were replaced with high-frequency synonyms so that EFL learners could better comprehend the texts. There were two versions of the passages: inference and neutral.

The inference texts were divided into two sub-versions, predictive and disconfirming. The predictive texts consisted of four sentences and approximately 50 words. These texts were designed to strongly induce specific predictive inferences about the outcomes of the described events; the target inference is strongly constrained by the contexts and related to the motivation for a character’s action (Klin, Murray et al., 1999; Nahatame, 2012). For example, in the sample predictive text in Table 2, the target inference “the boy hits the ball” is strongly suggested by the last sentence of the text, and is activated to explain why the boy raised his bat.

In contrast, the disconfirming texts were newly constructed by adding a continuing sentence
(e.g., the italicized sentence in Table 2) to the predictive texts that disconfirms the target inference suggested by the predictive text. Each disconfirming sentence was composed of 15 to 20 simple words, and created so that it naturally follows the preceding context. The plausibility of these sentences had been validated in the pilot study of Nahatame (2012).

The neutral texts, similar in length to the predictive text, also consisted of four sentences. However, these texts described topics that were different than those described in the inference texts, and did not induce target inferences. These neutral texts were used to provide a baseline measure of inference activation. In total, three sets (predictive, disconfirming, and neutral) of 16 experimental passages were used.

Table 2 Sample of Experimental Passages

<table>
<thead>
<tr>
<th>Inference Text (Predictive and Disconfirming)</th>
<th>Neutral Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>The boys’ high school baseball team was having tests for the spring season. The coach decided to test the boys’ baseball skills before he did anything else. The first batter to step up to the plate was a new boy on the team. As the pitcher released the ball, the boy raised his bat and the ball went directly towards him. Suddenly, the ball dropped in front of the bat and fell in the catcher’s mitt.</td>
<td>Halloween was a dark time for the citizens of the town. Two young boys had disappeared the day before. The last time they were seen was in a store with their mother. The boys had been trying on Halloween costumes when they vanished.</td>
</tr>
<tr>
<td>Target sentence for the MJT: <strong>The boy hits the ball</strong></td>
<td>Target sentence for the MJT: <strong>The boy hits the ball</strong></td>
</tr>
</tbody>
</table>

Note. The italicized sentence appeared only in the disconfirming text.

Each experimental passage was paired with a target sentence for the MJT (MJT Target) and a simple comprehension question (CQ). Each MJT Target simply described the future event suggested by the predictive text, including a main action verb directly representing a target inference concept (e.g., *hit*). The target verbs had familiarity ratings greater than five on a 7-point scale (Yokokawa, 2006) \( M = 5.99, \ SD = 0.51 \), and appeared in the most frequent 2,000 word level (Levels 1 and 2) in the JACET 8000 (JACET, 2003) \( M = 1.06, \ SD = 0.25 \). All of the sentences were composed of four or five words. The number of letters \( M = 17.69, SD = 2.94 \) and syllables \( M = 5.31, SD = 1.14 \) were also controlled. Furthermore, these sentences were written in the present tense and active voice to avoid the effects of participants’ grammatical knowledge on the meaningfulness judgment. CQs were yes or no questions about an explicitly stated piece of text information (e.g., *Was the first batter a new boy on the team?*). These questions were included to ensure the participants engaged in the reading.

Sixteen filler passages were used to balance the MJT responses. These fillers were similar in length to the experimental passages, but they did not elicit an inference. All MJT Targets paired
with the filler passages were semantically unacceptable (e.g., *They drink the hamburger*), though they included words related to the text content.

### 2.2.2 Target sentences for sentence recognition task

Following Muramoto (2000), the target sentences for the SRT (SRT Targets) were created for each inference text. Each text was paired with three types of SRT Targets: (a) an explicit sentence, which described an event explicitly mentioned in the text; (b) an inference sentence, which described a future event suggested in the predictive text; and (c) an inconsistent sentence, which described an event not mentioned or suggested in the text. All of these targets were written in Japanese to ensure that participants’ surface text memory about word forms and sentence structures did not have an effect on recognition judgment. Table 3 shows an example of three target types for the inference text described in Table 2. Each target was presented with a 4-point scale of recognition judgment confidence (1 = low, 2 = relatively low, 3 = relatively high, and 4 = high). This confidence scale allows for a finer-grained analysis of the participants’ recognition data.

<table>
<thead>
<tr>
<th>Explicit</th>
<th>Inference</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>コーチは少年達の技能をテストすることにした</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[The coach decided to test the boys’ skills.]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>少年はボールを打った</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[The boy hit the ball.]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>少年はグループを壊した</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[The boy cleaned a baseball glove.]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 2.2.3 L2 reading proficiency test

To measure the participants’ L2 reading proficiency, a test was created based on the reading subsection of the Test in Practical English Proficiency (EIKEN). Considering the participants’ estimated proficiency, two passages were chosen from retired copies of EIKEN Grade Pre-1, two passages from EIKEN Grade 2, and one passage from EIKEN Grade Pre-2. Each passage included three to five multiple-choice questions; the total number of items was 20.

### 2.3 Procedure

Participants were tested individually during the experiment, which lasted 90 minutes. An L2 reading proficiency test, which lasted for 30 minutes, was completed prior to the main experimental session. The main session included on-line and off-line phases as described below.

*On-line phase.* Four presentation lists (A-1, B-1, C-1, and D-1), each of which contained 12 experimental passages and 12 filler passages, were created for the on-line phase. Participants were randomly assigned to one of four lists. This ensured that each version of each passage (i.e., predictive, disconfirming, and neutral) was presented to an equal number of the participants.

The on-line phase was computer-based, administered by SuperLab 4.5 software. Each trial began with the word “Ready?” presented in the center of the computer screen (see Figure 1).
When participants were ready to begin the trial, they pushed the “Yes” button on the response pad and the first sentence of the passage appeared. Participants read the passage at their own pace, sentence by sentence. When the “Yes” button was pushed, the sentence was removed and replaced with the next one. After the final sentence of the passage was read, pressing the “Yes” button resulted in the appearance of a warning signal (XXX) for 750 ms in the center of the screen. The MJT Target followed the warning signal, flanked by asterisks. Participants were required to determine whether the target sentence was semantically acceptable as quickly and accurately as possible, using a pair of “Yes” and “No” buttons. The correct response is “Yes” for the experimental passages, whereas it is “No” for the filler passages. Accuracy and response latencies were recorded. After responding, another signal (???) appeared for 500 ms, then was replaced by a CQ. Participants made their responses and received accuracy feedback. This trial sequence was repeated for each passage. The passages were presented in random order.

Figure 1. The sequence of events during trials in the on-line phase of the experiment.

Note that participants completed the practice session before the experimental session. In the practice session, they engaged in the MJT for 15 practice target sentences to ensure that they understood the task requirements. In addition, the participants followed the same procedure as was used in the experimental session for five practice passages.

Off-line phase. Following the on-line phase, the participants received one of four presentation lists (A-2, B-2, C-2, D-2) for the off-line phase. Each list included eight inference texts and four filler texts, all presented on paper. Each participant received the list, including the passages they had not read in the on-line phase. The participants were asked to read these texts in a self-paced manner.

After the participants finished reading all the passages, they were given a worksheet in which three SRT Targets were presented for each text. They then were required to decide (a)
whether the target sentence was written in the text they had just read, and (b) how confident they were of their judgment on a 4-point scale.

2.4 Scoring and Data Analysis

MJT response data were eliminated if participants mistakenly skipped a sentence in the passage, because skipping sentences might have affected their accuracy and latencies. In addition, response times greater than 2.5 standard deviations (SDs) above the mean for each participant were excluded. This resulted in the exclusion of 3.24% of the data set. Prior to the analyses, the response time data were normalized by using a logarithmic transformation.

To examine the suppression of predictive inferences, inference activation scores were calculated from the correct response times of the MJT. These scores were obtained by subtracting the mean response times for the inference texts (i.e., predictive and disconfirming) from those for the neutral texts (Virtue et al., 2006). For example, if the mean response time was 2,000 ms for the predictive texts and 2,200 ms for the neutral text, the activation score for the predictive condition was 200 ms. If readers activate target inferences during reading, correct responses to MJT Targets should be facilitated (i.e., primed) by the activation, resulting in activation scores greater than zero. In contrast, if the activation of inferences is suppressed after reading the disconfirming context, correct responses to MJT Targets should not be facilitated, resulting in lower activation scores.

The data from the SRT was analyzed to investigate the revision of predictive inferences. Based on Muramoto (2000), the recognition ratings were calculated from participants' responses to SRT Targets and their confidence level, as shown in Table 4. The important focus is the recognition ratings for inference sentences. If readers activated target inferences and then encoded the inferences as part of text memory, inference sentences are likely to be falsely recognized as written in the text (perhaps with a high confidence level), resulting in higher recognition ratings. On the other hand, if readers initially activated the inferences but later revised them in their text memory, inference sentences are likely to be correctly recognized as not written in the text (perhaps with a high confidence level), resulting in lower recognition ratings.

| Table 4 Recognition Ratings Calculated From the Response and Confidence Level of the SRT |
|--------------------------------------------|---------------------------------|--------------------------------|
| Response                                    | Confidence level    | Recognition ratings |
|--------------------------------------------|--------------------------------|
| This sentence was **written** in the text.  | 4 (high)              | 6                  |
|                                            | 3 (relatively high)     | 5                  |
|                                            | 2 (relatively low)       | 4                  |
|                                            | 1 (low)                 | 3                  |
| This sentence was **not written** in the text. | 1 (low)                 | 3                  |
|                                            | 2 (relatively low)       | 2                  |
|                                            | 3 (relatively high)      | 1                  |
|                                            | 4 (high)                | 0                  |
3. Results

3.1 L2 Reading Proficiency Test and Comprehension Questions

The reliability of the L2 reading proficiency test was sufficient (Cronbach’s α = .79) after excluding three low discriminability items. Before the main statistical analysis, a median split was used to assign the scores to either a higher proficiency group (n = 18, M = 13.53, SD = 1.36, Max = 17, Min = 12) and a lower proficiency group (n = 18, M = 8.33, SD = 2.59, Max = 11, Min = 3), t(24.44) = 8.01, p < .001, d = 2.67.

The overall average percent correct for the CQs in the on-line phase was quite high (M = 92.80%, SD = 8.22), supporting the claim that participants engaged in the text reading. A 3 (Text: Predictive, Disconfirming, Neutral) × 2 (Proficiency: Higher, Lower) analysis of variance (ANOVA) on the means for the experimental passages did not indicate significant main effects of Text and Proficiency, F(2, 68) = 1.14, p = .326, ηp² = .032; F(1, 34) = 1.25, p = .272, ηp² = .035; the interaction was not statistically significant, F(2, 68) = 0.46, p = .637, ηp² = .013. Thus, both groups of participants comprehended the three text types equally well.

3.2 Meaningfulness Judgment Task

For the MJT, the mean percent correct was 92.79% (SD = 4.83) for all passages including fillers, and 90.49% (SD = 6.94) for the experimental passages only. Table 5 shows the response times for the correct responses and error rates (%) in each text condition. Figure 2 represents inference activation scores calculated from the correct response times. Before examining the suppression of inferences, it is necessary to confirm that target inferences were activated when participants read the predictive text. Thus, two-tailed paired t tests were conducted on mean inference activation scores for the predictive condition to determine whether the activation scores were greater than zero. The activation scores for both the higher and lower proficiency readers and all readers were significantly greater than zero, t(17) = 3.23, p = .005, d = 0.74; t(17) = 2.39, p = .029, d = 0.55; t(35) = 3.85, p < .001, d = 0.63. Thus, it was confirmed that participants activated the expected inferences when they read the predictive context.

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Predictive</th>
<th>Disconfirming</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Higher (n = 18)</td>
<td>1793.30 (9.72)</td>
<td>390.13</td>
<td>1914.83 (6.02)</td>
</tr>
<tr>
<td>Lower (n = 18)</td>
<td>2222.64 (8.33)</td>
<td>476.39</td>
<td>2325.76 (11.11)</td>
</tr>
<tr>
<td>Total (N = 36)</td>
<td>2007.97 (9.03)</td>
<td>481.20</td>
<td>2120.29 (8.56)</td>
</tr>
</tbody>
</table>

To examine the suppression of the activated predictive inferences, a 2 (Text: Predictive, Disconfirming) × 2 (Proficiency: Higher, Lower) ANOVA was conducted on the mean inference
activation scores. If the activation of target inferences was suppressed after reading the disconfirming context, the scores should be significantly lower in the disconfirming condition than the predictive condition. However, the main effects of Text and Proficiency, and the Text × Proficiency interaction were not statistically significant, $F(1, 34) = 0.81, p = .375, \eta^2_p = .023$; $F(1, 34) = 0.11, p = .743, \eta^2_p = .003$; $F(1, 34) = 0.15, p = .705, \eta^2_p = .023$. Thus, there were no significant differences in activation scores between the predictive and disconfirming conditions regardless of participants’ L2 reading proficiency, suggesting that inference activation was maintained even after processing the disconfirming context.

3.3 Sentence Recognition Task

The mean recognition ratings for each sentence type and text condition are reported in Table 6 and Figure 3. To examine the revision of inferences, a 3 (Sentence: Explicit, Inference, Inconsistent) × 2 (Text: Predictive, Disconfirming) × 2 (Proficiency: Higher, Lower) ANOVA was conducted on the recognition ratings. If readers revised the target inferences after reading the disconfirming information, these ratings should be lower in the disconfirming condition than in the predictive condition, only for inference sentences. The ANOVA revealed significant main effects of Sentence and Text, $F(2, 68) = 168.23, p < .001, \eta^2_p = .832$; $F(2, 68) = 29.56, p < .001, \eta^2_p = .465$, and a significant Sentence × Text interaction, $F(2, 68) = 14.96, p < .001, \eta^2_p = .306$. None of the other effects were significant or marginally significant, all $ps > .10$.

| Table 6 Mean Recognition Ratings for the Sentence Recognition Task |
|-----------------|-----------------|-----------------|-----------------|
|                 | Explicit        | Inference       | Inconsistent    |
|                 | Predictive      | Disconfirming   | Predictive      | Disconfirming   | Predictive      | Disconfirming   |
|                 | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ |
| Higher          | 4.94 | 0.82 | 4.19 | 1.18 | 2.39 | 1.07 | 0.92 | 0.55 | 1.05 | 0.76 |
| Lower           | 5.06 | 0.75 | 3.97 | 1.40 | 2.85 | 1.31 | 1.28 | 1.03 | 1.11 | 0.85 |
| Total           | 5.00 | 0.78 | 4.08 | 1.28 | 2.62 | 1.20 | 1.10 | 0.83 | 1.08 | 0.80 |

Note. The ratings ranged from 0 to 6.

Follow-up tests of the significant interaction revealed that there were significant differences between all sentence types for both predictive and disconfirming conditions ($ps < .001, ds > 0.80$).
More importantly, the simple main effect of Text was significant for inference sentences ($p < .001$, $d = 1.18$), but not for explicit and inconsistent sentences ($p > .10$, $d < 0.20$). The ratings for inference sentences were significantly higher in the predictive condition than the disconfirming condition. Thus, target inferences were more likely to be correctly recognized as not written in the disconfirming texts compared to the predictive texts. Nevertheless, the mean ratings for the inference sentences in the disconfirming condition suggest that participants recognized these sentences as not written with a low to relatively low confidence level (see Tables 4 and 6 again).

4. Discussion

**Suppressing the activation of incorrect predictive inferences during reading (H1)**

The first hypothesis proposed that the activation of predictive inferences is not suppressed during EFL reading, even when the subsequent context disconfirms the inferences. The statistical analysis of the inference activation scores, calculated from the correct response times of the MJT, showed that both higher and lower proficiency readers initially activated target inferences when they read the predictive context. However, the activation of inferences was maintained, not suppressed, after processing the disconfirming context, as indicated by the lack of a significant difference in the activation scores of the text conditions. Thus, H1 was supported. These results are highly consistent with those of Nahatame (2012), who used a word verification task; this suggests that the findings are not likely the result of using a specific task. Therefore, converging evidence from two different tasks indicates that EFL readers have difficulty suppressing the activation of predictive inferences immediately after reading a disconfirming context.

It is important to consider why the EFL readers had trouble suppressing predictive inferences. Given that L2 readers require more cognitive resources than L1 readers for lower level
text processing, they might find the suppression of inferences difficult because they had few
cognitive resources available for achieving the suppression. Iseki (2006) used the MJT as in the
present study, and demonstrated that predictive inference activation was suppressed by L1 readers.
However, it should be noted that the experimental passages used in the present study were
relatively short and simple so that the learners could easily comprehend them. In addition,
inference activation was not suppressed, regardless of participants’ L2 proficiency level. Taken
together, although inference suppression might have been difficult because of the cognitive
demands of L2 reading, this finding is attributable to a more direct influence of other factors.

One such factor is the amount of text information disconfirming the inferences. Although
previous studies indicated that readers’ comprehension difficulties occur when earlier-stated text
information contradicts newly stated information, these difficulties can be reduced by increasing
the amount of information that encourages them to modify their comprehension (e.g., Guéraud,
Harmon, & Peracchi, 2005). In the present study, the information disconfirming the target
inference was described in just one sentence, whereas the context preceding inference activation
consisted of four sentences. Although the context indicated that the activated inferences were
incorrect, it seemed as though the amount of context was insufficient to fully suppress the inferences.

Another factor is the characteristics of the target inferences; these inferences were all
activated to explain the motivation for the characters’ actions (e.g., The boy raised his bat because
he tried to hit the ball). Thus, a coherent comprehension of the passage required a strong
activation of target inferences, which might make it difficult to immediately suppress the
inferences even after the disconfirming information. In addition, although the inferred future event
(e.g., The boy hits the ball) was disconfirmed by the subsequent context, the character’s
motivation for the event (e.g., The boy tried to hit the ball) was not actually disconfirmed. In this
case, the immediate suppression of the target inferences might not be highly necessary.

However, even if the suppression of inferences was not immediately required, readers need
to understand that the inferred event did not actually happen in the described situation. If they
falsely maintain disconfirmed inferences in text memory without revision, they may fail to
accurately comprehend the text. The next section discusses this issue.

Revising incorrect predictive inferences after reading (H2)

The second hypothesis was that disconfirmed predictive inferences are not revised in EFL
readers’ text memory. The implication of the SRT results differ from that of the recall task results
of Nahatame (2012). The statistical analysis of the recognition ratings revealed significant
differences between the predictive and disconfirming conditions only for inference sentences, with
higher ratings in the predictive condition than in the disconfirming condition. This suggests that
readers revised the target predictive inferences after they read the disconfirming context.

However, it should be noted that the recognition ratings in the disconfirming condition were
significantly higher for inference than inconsistent sentences. Furthermore, the mean recognition
ratings suggested that the participants were less confident in their judgment when they rejected the inference sentences in the disconfirming condition. Taken together, the most plausible interpretation of the SRT results is that the participants partially, but not completely, revised the disconfirmed inferences. In other words, although EFL readers tend to revise the predictive information after the context disconfirms it, it can be difficult to eliminate the memory trace of the strongly predicted information (H2 was partially supported). In this regard, compared to the findings of Nahatame (2012), the findings of the present study provided a more accurate picture of EFL readers’ revisions of predictive inferences.

When the pattern of results for the MJT is compared to that of the SRT, there appears to be an inconsistency. The EFL readers found it difficult to suppress their inferences during reading, but they partially revised predictive inferences after reading. Although the text representations during reading and after reading should be distinguished, it may be important to note that the participants’ reading experience before the MJT was different than that before the SRT. Specifically, before the MJT, they read all of the passages sentence by sentence. Before the SRT, they read all of the passages with all sentences presented at once. Thus, the difference in the pattern of results of the two tasks might be caused by whether readers could read the earlier context when the disconfirming context is encountered; in other words, a reanalysis of activated inferences based on the earlier context may play a role in suppressing and revising incorrect inferences.

5. Conclusion

This study investigated EFL learners’ flexible text comprehension processes, focusing on predictive inferences in reading. Specifically, it examined the suppression and revision of incorrect predictive inferences with an awareness of methodological issues in evaluating learners’ inference activation and text memory. The results of this study confirmed and extended earlier findings. First, consistent with Nahatame (2012), the results of the MJT suggested that EFL readers have difficulty suppressing the activation of predictive inferences immediately after reading the context disconfirming the inferences. In contrast, the SRT results revealed that they achieved a partial revision of inferences in text memory based on the disconfirming context.

With these findings in mind, it is helpful to be cognizant of the limitations of this study. First, this study included participants with an approximately intermediate level of L2 proficiency. Thus, although there was a lack of an effect of L2 proficiency on suppression and revision, this lack might be due to the limited range of their proficiency. Further research is required involving participants with more L2 proficiency variations. Second, although the present study suggested a few factors that likely affected participants’ difficulty in suppressing inferences, it was difficult to identify which was the determinant factor. Therefore, it is necessary to directly examine the effects of these factors such as the amount of disconfirming information, characteristics of
inferences, and text presentation mode. Finally, it is important to investigate in detail how learners process the context disconfirming the inferences. This issue can be addressed by examining learners’ think-aloud protocols or eye movements during reading. Such research will be valuable because it provides information that learners can use to achieve the suppression of incorrect inferences, leading to effective and flexible control of their attention resources in reading.

As described in the Introduction, learners make predictions about what might be described in the subsequent context, both spontaneously and strategically. However, when learners make strong predictions that are immediately disconfirmed in the following context, they may have trouble suppressing the activation of the predictions during reading, and may only partially revise the predictions in memory. This is an important issue that teachers should bear in mind, and they should help students more successfully suppress and revise their predictions in such a case. In this regard, it might be helpful to give students some questions concerning the predictions (e.g., What happened to the batter?) and ask the students to provide reasons for their answers, rather than simply asking them to translate the disconfirming context. This might facilitate a careful reanalysis and deeper understanding of the text. This study emphasizes the importance of considering such an approach to develop learners’ flexible comprehension processes in reading.

Notes

1. For example, The vase was broken is an acceptable sentence; The vase was cried is an unacceptable sentence. Meaningfulness judgment (also called acceptability judgment) is similar to grammaticality judgment; however, it is a broader concept (Jiang, 2012).

2. The error rates were low in each condition, and any significant effects were not found in a 3 (Text: Predictive, Disconfirming, Neutral) × 2 (Proficiency: Higher, Lower) ANOVA on error rates. Therefore, these data indicate that there was no speed-accuracy trade-off in the MJT.

Acknowledgements

I would like to thank Prof. Yuji USHIRO for his enormous support. I would also wish to express my appreciation to all three anonymous reviewers for their valuable comments. This research was supported by a Grant-in-Aid for JSPS Fellows (No. 24 • 1337).

References


