Making and Revising Predictive Inferences in Japanese EFL Learners' Reading Comprehension

A Dissertation
Submitted to the University of Tsukuba
In Partial Fulfillment of the Requirements for
the Degree of
Doctor of Philosophy in Linguistics

Shingo NAHATAME

2014

Abstract of the Dissertation

Making and Revising Predictive Inferences
in Japanese EFL Learners' Reading Comprehension
by

Shingo NAHATAME

One of the goals for students learning English as a foreign language (EFL) is to successfully read and comprehend texts written in English. For students to achieve this goal, teachers need to understand how their students read texts and what kind of difficulties they encounter in reading.

To date, studies have investigated the cognitive process of reading. In particular, inferential processing in reading has received considerable attention from researchers in psycholinguistics and from educators as well. Of several types of inferences produced in reading, *predictive inference*, which is the anticipation of the likely outcome of an event described in a text, has been widely investigated. Previous studies have provided valuable insights into whether and how predictive inferences are made in first language (L1) reading (e.g., Calvo & Castillo, 1996; Lassonde & O'Brien, 2009; McKoon & Ratcliff, 1986).

In contrast, inference generation in second language (L2) reading has been investigated by only a limited number of studies (e.g., Horiba, 1996, 2000; Muramoto, 2000; Yoshida, 2003). In addition, there are few studies with a special focus on predictive inference generation in L2/EFL reading. Because making predictive inferences contributes to fluent reading and deep text comprehension (Fincher-Kiefer, 1993; Linderholm, 2002),

investigating these inferences in L2/EFL reading will provide suggestions for improving learners' reading comprehension skills.

Although making inferences plays an important role in reading comprehension, readers are required to revise their initial inferences when their inferences are disconfirmed by the following context. However, few studies have investigated whether and how L2/EFL readers revise inferential comprehension. Thus, insights into the revision of predictive inferences in L2/EFL reading will be beneficial for developing learners' flexible text comprehension processes, which are required for learners to be successful readers (Ushiro, 2010).

To address the aforementioned issues and provide insightful implications for EFL reading instruction, two empirical studies were conducted within the current dissertation research, which investigated making predictive inferences (Study 1) and revising predictive inferences (Study 2) among Japanese EFL readers. Each study consisted of three individual experiments.

Experiment 1 examined the relationship between predictive inference generation and text characteristics in EFL reading. Japanese university students read several short narrative passages that were designed to elicit specific predictive inferences. These passages were also manipulated in terms of inference subtypes (i.e., whether predictive inferences relate to narrative characters' motivation) and contextual constraint (i.e., whether the context strongly constrains possible inferences). Immediately after reading each passage, the participants performed a recognition task to target words that are related to the expected inference. In addition, the participants engaged in a written recall task after reading all the passages. Recognition times for the target words and recall rates of inferential information indicated that the readers were most likely to make predictive inferences when the inferences were related to the character's motivation and strongly constrained by context.

Experiment 2 investigated the relationship between predictive inference generation and the amount of cognitive resources available in reading. Participants read short narrative passages under three different load conditions (i.e., zero-, low-, and high-load). Immediately after reading each passage, the participants performed a lexical decision task to target inference words. Additionally, the participants engaged in a probability judgment task for inference sentences after reading all the passages. Lexical decision times to target words suggested that predictive inferences were likely to be impaired when the amount of available cognitive resources was reduced. Moreover, the results of probability judgments raised the possibility that increased demands on cognitive resources specifically reduced the automaticity or immediacy of predictive inference generation in reading.

Experiment 3 explored the relationship between predictive inference generation and strategic text processing. In this experiment, participants were instructed to either understand the passages or anticipate the outcome of the events described (i.e., they were given strategy instructions). The participants engaged in a lexical decision task to target words and a written recall task as in previous experiments. Lexical decision times to target words revealed the facilitation effects of strategy instructions on predictive inference generation, and these effects were much larger among the proficient readers than among the less proficient ones. In addition, the analysis of recall productions showed that giving strategy instructions aimed at predictive inferences did not impair comprehension of explicit text information.

Experiment 4 investigated whether Japanese EFL readers revise predictive inferences disconfirmed by the following context. In this experiment, Japanese university students read short narratives designed to elicit predictive inferences (i.e., predictive texts) or to disconfirm the induced inferences (i.e., disconfirming texts). The participants engaged in the recognition and recall tasks similar to those used in Experiment 1. Recognition times to target words suggested that the readers failed to suppress the activation of inferences immediately after the disconfirming context. Similarly, recall rates of inferential information suggested that disconfirmed inferences were not deleted from the readers' long-term text memory. These trends were consistent between the proficient and less proficient readers.

To confirm the findings of Experiment 4, Experiment 5 was conducted using tasks different from those used in Experiment 4. Specifically, instead of the word-probe recognition task, participants completed a meaningfulness judgment task in which target sentences described the events predicted. In addition, after reading all the passages, the participants engaged in a recognition task to the target sentences, instead of the written recall task. Consistent with Experiment 4, judgment times confirmed that the readers had difficulty suppressing the activation of inferences immediately after the disconfirming context, regardless of L2 proficiency. In contrast, recognition performance suggested that the readers partially deleted the disconfirmed inferences from text memory, but it was still difficult to achieve the complete deletion.

Experiment 6 investigated readers' real-time comprehension processes when they encounter a context disconfirming predictive inferences. In this experiment, the participants' eye movements during reading were recorded. The analysis of eye movements on the disconfirming context suggested that both the proficient and less proficient readers consciously noticed the inconsistencies between the drawn inferences and the disconfirming context during text processing; however, the less proficient readers experienced more difficulty in integrating the disconfirming context into developing text representations.

Study 1 of this dissertation revealed the conditions in which Japanese EFL learners are more or less likely to make predictive inferences while reading. This study concluded that predictive inference generation in EFL reading is a more limited and complicated process than in L1 reading as a result of interaction between text, reader, and task factors. Additionally, Study 2 indicated that Japanese EFL learners have difficulty revising predictive inferences disconfirmed by the immediately following context. Importantly, this study suggested the difficulty specific to less proficient learners' reading process: integration of the context disconfirming the inferences into developing text representations during text processing.

The current dissertation provides insights into whether and how Japanese EFL learners make and revise predictive inferences in reading. The findings of the present studies have a number of educational implications for focusing leaners' attention on making predictive inferences and developing their flexible comprehension processes in reading.

Acknowledgements

I really appreciate many people for their insightful guidance and kind support during the course of writing the current doctoral dissertation at the Graduate School of Humanities and Social Sciences, University of Tsukuba. Although I appreciate everyone who has assisted me in writing this dissertation, the following people deserve a special mention.

My deepest appreciation goes to my academic supervisor, Professor Yuji Ushiro at the University of Tsukuba. He awakened my interest in the second language learners' reading process and inspired me to learn more about it. I learned a great deal about this field from his class and his Reading Research Group at the University of Tsukuba. His comments and suggestions were always of inestimable value for my research. He encouraged me to pursue my goal in many ways throughout the course of my study.

I would also like to thank Professor Hirosada Iwasaki at the University of Tsukuba. His generous support and insightful advice were invaluable. In his class, he taught me much about how to write academic papers more effectively and properly. I am sure that my academic writing skills improved thanks to him. I am also grateful to Professor Akiyo Hirai at the University of Tsukuba. She gave me a number of invaluable comments about experimental design and procedures of my research. In addition, she provided me with important knowledge on how to analyze and interpret data in experimental studies in her class. This knowledge definitely helped me to conduct the current research. I am also indebted to Professor Shuhei Kadota at the Kwansei Gakuin University, who willingly accepted the invitation to be an external member of my dissertation committee. He provided insightful comments on my research in terms of psycholinguistics and second language acquisition. They greatly improved an earlier version of this dissertation.

I have greatly benefited from the help of senior and junior associates in Professor

Ushiro's seminar. In particular, I would like to express my gratitude to Ms. Haruka Shimizu, Ms. Yukino Kimura, and Mr. Akira Hamada for their constructive comments and generous support. I would also like to express my heartfelt thanks to my colleague, Mr. Yusuke Hasegawa. He has constantly supported me throughout this study, and the observations he has made have always been insightful.

I also owe a very important debt to the 212 students at the University of Tsukuba who participated in my experiments. The current research could not have been conducted without their kind cooperation.

I would like to offer my sincere thanks to my dear friends for their constant and warm encouragement. Their generous and heartfelt words certainly helped me while I was writing this dissertation.

Finally, I would like to express my profound gratitude to my parents for their moral support and sincere encouragement. My father understood my eagerness to study, and my mother always looked upon me warmly.

Shingo NAHATAME

Tsukuba, Japan

December, 2014

Table of Contents

Chapter 4 Study 2: Revising Predictive Inferences in Japanese EFL Learners' Reading Comprehension

List of Tables

Chapter 4

List of Figures

List of Appendices

Major Abbreviations and Acronyms

L1 = First Language

L2 = Second Language

EFL = English as a Foreign Language

C-I model = Construction-Integration model

SBF = Structure Building Framework

E-I model = Event Index model

WM = Working Memory

MJT = Meaningfulness Judgment Task

SRT = Sentence Recognition Task

MSC = Motivational-Strong Constraint

MWC = Motivational-Weak Constraint

CSC = Consequence- Strong Constraint

CWC = Consequence-Weak Constraint

ANOVA = Analysis of variance

MANOVA = Multivariate analysis of variance

Chapter 1

Introduction

1.1 Background of the Current Research

One of the crucial goals for students learning English as a foreign language (EFL) is to successfully read and comprehend a text. Although the basic text comprehension processes are not different between first language (L1) and second language (L2) reading, the complexity and difficulty of the processes obviously increase in L2 reading (Grabe, 2009; Koda, 2005). Therefore, a clear understanding of how EFL learners read a text and the difficulties they encounter in reading should be beneficial for EFL classroom instruction.

To date, many studies have been carried out on how readers comprehend a text using several cognitive abilities (e.g., memory, problem-solving, and inferences). In particular, readers' inferential processing has received attention from numerous researchers in psycholinguistics and from educators as well. This is because inference making (i.e., comprehension of an implicit idea strongly suggested in a passage) plays a significant role in reading comprehension by establishing textual coherence or embellishing what is explicitly stated in the text. As inference generation is of critical importance in reading, the ability to make appropriate inferences about a text is often contained in syllabi. Moreover, it is measured in formal language tests for both L1 and L2 reading. For example, academic reading skills measured by the TOEFL iBT® include "making inferences about what is implied in a passage" (Educational Testing Service, 2012, p. 7).

Over the past 35 years, studies have provided valuable insights into how and what kinds of inferences are made in L1 reading (e.g., Graesser, Singer, & Trabasso, 1994; van den Broek, 1994; van Dijk & Kintsch, 1983). Of the group of possible inferences that may occur during reading, many studies have given special attention to *predictive inference*, which is the anticipation of the likely outcome of an event described in a narrative text (e.g., Calvo & Castillo, 1996; Lassonde & O'Brien, 2009; McKoon & Ratcliff, 1986). Predictive inferences benefit readers in ways such as easing the processing of the subsequent context, promoting construction of a situation model (the ideal form of text representations), and encouraging active engagement with the text (Allbritton, 2004; Fincher-Kiefer, 1993; Linderholm, 2002). As a result of a growing body of research, some aspects of predictive inference generation in L1 reading have been revealed.

In the area of L2 reading, researchers have also investigated whether L2 learners make inferences in reading, and if so, what types of inferences they make (e.g., Horiba, 1996, 2000; Muramoto, 2000; Yoshida, 2003). However, as Koda (2005) pointed out, the number of studies on inference generation in L2 reading is limited. Additionally, these studies have not closely looked into particular types of inferences in the EFL context. Therefore, it remains uncertain whether and how specific types of inferences are made in EFL reading. For this reason, the current research was carried out with a focus on predictive inferences, which have been widely investigated in L1 reading research, in Japanese EFL learners' reading comprehension.

Although making inferences plays an important role in reading, readers do not always make correct inferences. Sometimes, when they read the subsequent text, they find that their inferences were incorrect. In this case, it is necessary to revise their initial inferences in order to achieve appropriate comprehension of the text. Ushiro (2010) asserted that active and flexible revision of text comprehension is key for EFL learners to become successful readers. Despite its importance, however, few studies have investigated whether and how EFL learners revise inferential comprehension in reading. Thus, the current research also aims to examine the processes of revising predictive inferences among Japanese EFL readers.

In sum, the objective of the current research is to examine the following two reading processes among Japanese EFL learners: *making* and *revising predictive inferences*. Insights into these processes will reveal important aspects of EFL learners' text comprehension mechanism and possible difficulties occurring in EFL reading. The findings will also have educational implications for developing effective and flexible reading comprehension processes among Japanese EFL learners.

1.2 Organization of This Dissertation

The current dissertation consists of the following five chapters: Introduction (Chapter 1; the current chapter), Review of Related Literature (Chapter 2), Study 1 (Chapter 3), Study 2 (Chapter 4), and General Discussion and Conclusion (Chapter 5).

Chapter 2 provides a review of previous studies related to the current research. First, basic concepts of reading comprehension are introduced, along with major theoretical models proposed by L1 studies. Second, theories of inference generation during reading are described as well as empirical studies on inference generation in L2 reading. Third, previous studies on predictive inferences in reading are carefully examined, focusing on factors affecting the generation and measurement of the generation. This chapter concludes with some issues in previous studies that remain unresolved and need to be addressed in the current research.

The current dissertation includes two empirical studies, each of which consists of three experiments. The overview of these two studies is illustrated in Figure 1.1. Study 1 investigated the generation of predictive inferences, whereas Study 2 examined the revision of the inferences in Japanese EFL learners' reading comprehension.

Study 1: Making Predictive Inferences in Japanese EFL Learners' Reading Comprehension		Factors	On-line Task	Off-line Task
Experiment 1	Predictive inference generation and text characteristics in EFL reading	Contextual Constraint Inference Subtypes	Recognition	Written Recall
Experiment 2	Predictive inference generation and cognitive demands in EFL reading	Cognitive Demands (zero, low, high)	Lexical Decision	Probability Judgment
Experiment 3	Predictive inference generation and strategic processing in EFL reading	Instruction Types L2 Reading Proficiency	Lexical Decision	Written Recall

Study 2: Revising Predictive Inferences in Japanese EFL Learners' Reading Comprehension		Factors	On-line Task	Off-line Task
Experiment 4	Examining the revision of predictive inferences in EFL reading	Text Types (predictive, disconfirming) L2 Reading Proficiency	Recognition	Written Recall
Experiment 5	Reexamining the revision of predictive inferences in EFL reading	Text Types (predictive, disconfirming) L2 Reading Proficiency	Meaningfulness Judgment	Sentence Recognition
Experiment 6	Revision of predictive inferences and text comprehension processes in EFL reading	Text Types (disconfirming, control)	None (Eye tracking)	Sentence Recognition

Figure 1.1. Overview of the two experimental studies in the current dissertation.

Chapter 3 reports on Study 1, which investigated predictive inference generation in EFL reading. Experiment 1 examined the relationship between predictive inference generation and text characteristics. In this study, Japanese university students read a set of short narratives designed to elicit predictions of likely outcomes of events described therein. These passages were manipulated in terms of two critical text factors: contextual constraint (i.e., whether context strongly constrains possible inferences) and inference subtypes (i.e., whether inferences relate to the motivation for narrative characters' actions). Immediately after reading each passage, participants performed a recognition task to target probe words that are related to the expected inference. In addition, participants engaged in a written recall task after reading all passages. Recognition times for target words and recall rates of inferential information were analyzed to examine the effects of text characteristics on predictive inference generation in reading.

Experiment 2 investigated the relationship between predictive inference generation and cognitive demands in reading. This experiment constructed three experimental conditions, across which the degree of cognitive demands given to participants varied (i.e., zero-, low-, and high-load conditions). Immediately after reading each passage, participants performed a lexical decision task to target words, and engaged in a probability judgment task to target sentences after reading all passages. In the low- and high-load conditions, participants were also asked to retain a list of words in their memory while reading the passages. Lexical decision times and probability ratings were analyzed to examine the effects of cognitive demands on predictive inference generation in reading.

Experiment 3 explored the relationship between predictive inference generation and strategic text processing. In this experiment, before reading the several narratives, participants were instructed to either understand the passages or anticipate the outcome of the events described. Each passage was followed by a lexical-decision probe word, and a written recall task was conducted after reading all passages. Lexical decision times were analyzed to investigate the effects of instruction type on predictive inference generation, whereas recall performance was analyzed to examine these effects on comprehension of explicit text information. The effects of instruction type were also examined in term of leaners' L2 reading proficiency.

Chapter 4 describes Study 2, which examined the revision of predictive inferences in EFL reading. Experiment 4 investigated whether Japanese university students suppress the activation of predictive inferences after the subsequent context disconfirms the inferences, and whether the students delete the disconfirmed inferences from long-term text memory. In this experiment, participants read short narratives designed to elicit predictive inferences (i.e., predictive texts) or to disconfirm the induced inferences (i.e., disconfirming texts). Participants also engaged in the recognition and recall tasks used in Experiment 1. Recognition times and recall performance were compared between predictive and disconfirming text conditions, including consideration of L2 reading proficiency.

To confirm the findings of Experiment 4, Experiment 5 was conducted using tasks different from those used in Experiment 4. Specifically, instead of the recognition task, participants completed a meaningfulness judgment task in which target sentences described the events predicted from the passages. In addition, after reading all passages participants engaged in a recognition task to target sentences, instead of the written recall task. Response times for the judgment task and recognition performance were compared between predictive and disconfirming text conditions, including L2 reading proficiency as a reader factor.

Experiment 6 investigated readers' text comprehension processes when they encounter the context disconfirming predictive inferences. Participants read disconfirming texts or control texts—in which no disconfirmation of inferences was included—and their eye movements during reading were recorded. After reading all the passages, the same recognition task as in Experiment 5 was conducted to confirm the findings of Experiment 5. Readers' eye movements were analyzed to investigate the kinds of reading processes that are more related to the difficulty of revising predictive inferences.

Chapter 5 generally discusses the findings of these experiments and draws conclusions about making and revising predictive inferences in Japanese EFL learners' reading comprehension. The limitations of the current research are also described, along with suggestions for future research. Finally, this chapter proposes educational implications of the present findings for reading instruction in EFL classrooms.

Chapter 2

Review of Related Literature

2.1 Reading Comprehension

2.1.1 Models of Reading Comprehension

To date, several models of reading comprehension have been proposed by researchers. This section briefly introduces three specific models related to the current research: the construction-integration model (Kinstch, 1988), structure building framework (Gernsbacher, 1990), and event-indexing model (Zwaan, Langston, & Graesser, 1995). Several concepts associated with these models and some empirical studies are also described.

The construction-integration model

The construction-integration (C-I) model was first introduced by Kintsch (1988), although its foundations can be traced to earlier studies (Kintsch & van Dijk, 1978; van Dijk & Kinstch, 1983). Van Dijk and Kinstch (1983) proposed three levels of mental representations readers construct from the text: the *surface form*, *textbase*, and *situation model*. At the first level, the surface form represents the actual order of words written in the text. At the second level, the textbase is a representation of the propositional network extracted from a text. A textbase is limited in that it is the representation of what is described in the text without considering readers' general knowledge. The third level, the situation model is the most enduring level of representation, and represents the content that the text is about. The construction of situation models requires readers to combine the textbase with their general knowledge from long-term memory through active inference generation. Although there are some occasions when reading requires only the construction of a textbase, it is the construction of the situation model that indicates readers' deeper and richer understanding of a text.

Kintsch (1988) developed his own idea and introduced the C-I model. According to this model, the construction of a text comprehension proceeds in cycles of two steps. The first step is to construct a network of propositional units in a text and their interrelations (i.e., construction phase). These construction processes need not be precise, and created networks sometimes include irrelevant or contradictory elements. To get rid of these irrelevant and contradictory elements, readers engage in the second step (i.e., the integration phase). The integration processes are conceptualized as spreading activation through the propositions in the network. As a result of the integration processes, the relevant propositions receive strong activations, whereas irrelevant ones become deactivated.

The C-I model was further developed in a series of studies by Kintsch (1998) as a way to clarify the nature of comprehension. Kintsch, however, primarily emphasized the distinction between textbase and situation models of text representations, as previously described by van Dijk and Kintsch (1983). Indeed, a number of subsequent studies provided evidence for the distinction between the different levels of the text representations.

McNamara, Kintsch, Songer, and Kintsch (1996), for example, indicated the significant interaction among readers' knowledge about the text topic, textual coherence, and levels of representations. They demonstrated that knowledgeable readers developed different levels of representations according to the coherence of the text. In this study, readers with sufficient knowledge about the text topic gained higher scores on the post-reading questions assessing situational understanding of the text, such as the inference and problem-solving questions, when reading the incoherent text compared to the coherent text. On the other hand, their scores on the post-reading questions assessing superficial understanding of the text, such as text-based questions, showed the opposite pattern; that is, the scores were higher when reading a coherent text than when reading an incoherent text. These results suggest that knowledgeable readers constructed situation models from the incoherent text through active inference generation based on their background knowledge, whereas they did not engage in text reading when reading the coherent text, resulting in the formation of textbase representations. This finding supports the view of multilevel text representations readers construct, suggesting that the textbase and the situation model should be distinguished from each other.

The structure building framework

The structure building framework (SBF) is an interesting complement to Kintsch's (1988) C-I model. This model explains how discourse comprehension is constructed through sentence-by-sentence processing (Gernsbacher, 1990, 1997). The SBF assumes that to achieve coherent comprehension, readers must first lay the foundation for mental representations. Indeed, studies have shown that readers slow down when reading the initial sentences of each paragraph or episode, suggesting that they use these sentences to lay the foundation for mental structures representing paragraphs and episodes.

Then readers develop their mental representations using two general cognitive processes: *mapping* and *shifting*. When the incoming information is related to the previous information, it is mapped onto the representations of previously comprehended information. On the other hand, if the incoming information is less related, readers shift and develop a new substructure of the representations. These two processes are achieved by two cognitive mechanisms modulating the activation of concepts: *enhancement*, which increases the activation of relevant concepts in the representation, and *suppression*, which decreases the activation of less relevant concepts in the representation. Although both the SBF and the C-I model proposed that readers activate information maintained in their memory when it is related to the information currently being processed, these two theories are different in their views of how readers deal with irrelevant information in their representations. The C-I model supposes the deactivation of irrelevant information based on an automatic inhibition process, whereas the SBF suggests active suppression of irrelevant information based on the suppression mechanism (Tapiero, 2007).

The results obtained in some previous studies provided the evidence for the suppression mechanisms suggested in the SBF. Gernsbacher, Robertson, Palladino, and Werner (2004) investigated the activation of a narrative main character during reading. The analysis of recognition times for target probes indicated that readers increased the activation of the main character's name or an object associated with the main character when the character was rementioned in the text, whereas the activation was suppressed when a new character was introduced. Similarly, Linderholm et al. (2004) confirmed that readers enhanced the activation of a narrative character's goal information when it was rementioned but suppressed it when a new goal was introduced. These results are consistent with the account provided by the SBF, and showed that this framework can be applied to both concrete types of text information, such as story characters' names, and abstract types of information, such as story characters' goals.

In addition, the SBF assumes that persons who are more skilled at comprehension are characterized as having a more efficient suppression mechanism. Interestingly, Gernsbacher et al. (2004) demonstrated that less skilled readers found it difficult to suppress the activation of the first-mentioned character's name after a new character was introduced, while more skilled readers successfully suppressed the activation of that information. These results support the relationship between comprehension skills and the suppression mechanism proposed by the SBF.

The event-indexing model

Zwaan, Langston, et al. (1995) proposed the event-indexing (E-I) model that accounts for the reader's construction of a situation model while reading. This model proposes that readers keep track of at least five situational dimensions during comprehension: *temporality*, *spatiality*, *protagonist*, *causality*, and *intentionality*. This model assumes that an event which took place at the same time and place as the previous event (i.e., these two events are continuous) will be more easily integrated into the developing situation models than an event that occurred at a different time and place. The studies of Zwaan and his colleagues (e.g., Zwaan, 1996; Zwaan, Magliano, & Graesser, 1995) demonstrated that a discontinuity of the situational dimensions significantly increased reading times for a sentence. For example, Zwaan (1996) compared the reading times of sentences containing a small temporal shift (e.g., a moment later) with those containing a large temporal shift (e.g., a day later). The result showed that the processing time of the former sentences was significantly shorter than that of the latter sentences.

In addition, Zwaan and Radvansky (1998) claimed that the situation model is divided into three types: (a) *the current model*, which is constructed while a person reads a particular clause or sentence, (b) *the integrated model*, which is constructed while the person integrates the upcoming information with prior information, and (c) *the complete model*, which is stored in long-term memory after all the textual input has been processed. The process of incorporating the current model into the integrated model is called *updating*.

A large body of research has been conducted so far to investigate the updating of situation models during reading. O'Brien, Rizzella, Albrecht, and Halleran (1998) examined whether the updating of situation models is completed by adding qualifying information to the text describing two inconsistent characteristics of a protagonist. For example, readers encounter the information about Mary being vegetarian in an earlier part of the passage (i.e., old information); however, this information is qualified with the subsequent information that Mary is not strict about her diet when she eats out. Then, readers encounter the target sentence that conflicts with the old information; *Mary ordered a cheese burger and fries* (i.e., new information). Analyzing the reading times for the target sentences through all five experiments, O'Brien et al. found that the reading times were longer when a text included an inconsistency than when it did not, even though the inconsistency was qualified with the additional information. Based on the results, O'Brien et al. put forward the view that old information is reactivated when readers encounter a sentence contradicting the information, and this reactivation interferes with the integration of the current information included in the contradicting sentence information (called *memory-based view*). Guéraud, Harmon, and Peracchi (2005) conducted similar experiments with O'Brien et al., but the amount of qualifying information was increased. They found that the additional qualifying information decreased comprehension difficulty caused by inconsistent information, but readers still reactivated the old information, supporting the view suggested in O'Brien et al.

Zwaan and Madden (2004) conducted some experiments with similar materials and procedures as O'Brien et al. (1998), but controlled the plausibility of the qualification included in the text. Contrary to the results obtained in O'Brien et al., they found that irrelevant and old information did not affect the processing of new information. Based on the results, they proposed the view that the most current information was more available than old and outdated information because readers always maintain the updated situation models (called *here-and-now view*). Other studies have also shown that readers successfully update their text comprehension if the qualification context provides a sufficient and plausible explanation for updating, or if pre-reading instructions encourage readers to carefully track the unfolding plot (Kendeou, Smith, & O'Brien, 2013; Rapp & Kendeou, 2007, 2009).

Long and Chong (2001) pointed out the difference between good and poor readers in the processing of inconsistent statements included in the text. They used passages where the inconsistent statements were either adjacent (i.e., local condition) or were separated by several sentences (i.e., global condition). The analysis of reading times for target sentences suggested that both good and poor readers showed inconsistency effects in the local condition, but in the global condition, only good readers demonstrated the effects. The following sentence verification task indicated that both good and poor readers reactivated earlier-stated information, but poor readers had more difficulty than good readers in integrating the information with the incoming information.

Some researchers have explored how readers deal with inconsistent statements during reading using eye-tracking. For instance, Rinck, Gamez, Diaz, and de Vega (2003) measured eye movements of L1 readers while reading a narrative passage containing temporally inconsistent information. For example, the sixth sentence of the passage presented critical information (e.g., *Claudia was already waiting for Mark when he arrived*), which was either consistent or inconsistent with information described in the second sentence (e.g., *Claudia arrived before Mark* / *Mark arrived before Claudia*). The results showed that eye movements on the sixth sentence were not affected by inconsistencies in the text; instead, inconsistencies caused more and longer regressions towards the previously stated information (i.e., the second sentence). Similarly, van der Schoot, Reijntjes, and van Lieshout (2012) examined how inconsistent character information is processed in reading of primary school children. The results were somewhat different from Rinck et al.: inconsistencies increased the duration of first-pass fixations on target sentences as well as total fixation durations at the end of the sentence. However, readers did not frequently make regressions towards previously stated information in this study.

As described above, researchers have provided conflicting accounts about whether and how old (or outdated) information continues to influence comprehension of new (or current) information. There is not yet a conclusive account, and researchers are still investigating this issue.

2.1.2 Characteristics of Narrative Reading

Because the current research focused on reading narrative texts, it is important to identify the characteristics of narrative reading. Several studies have attempted to identify the structural properties of narrative discourse. The following two approaches, *story grammar* (Mandler & Johnson, 1977; Stein & Glenn, 1979) and *causal network analysis* (Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985), have been widely used in narrative comprehension studies.

Story grammar

Story grammar focuses on the typical order in which several types of story events occur. Although story grammars vary between studies, the grammars typically categorize story events into the following components: *setting*, *initiating event*, *reaction*, *goal*, *action*, *outcome*, and *ending*. Stories usually begin with a setting that includes an introduction to the protagonists and their location in time and space. An initiating event is an event that provokes a reaction from the protagonist. The protagonist's reaction establishes his/her goal or motivation. The protagonist takes one or more intentional actions to achieve this goal. The actions result in an outcome that is either successful or unsuccessful in terms of the goal. The outcome concludes the episode and elicits an ending. The ending contains the protagonist's final reactions.

Story grammars predict which category is more crucial for comprehension and more memorable. Indeed, some studies have demonstrated that L1 comprehenders are sensitive to structures described by story grammars. They recall initiating events, goals, and outcomes more frequently and rate them as more important than reactions, endings, and actions (Mandler & Johnson, 1977; Stein & Glenn, 1979).

In addition, the grammars provide a hierarchical structure for describing story events presented in the text. Story episodes are often hierarchically related, such that the character's superordinate goal is not achieved and then a subordinate goal is generated. Story grammars predict that superordinate information in the hierarchy is more important and more memorable than subordinate information. A number of studies have confirmed this prediction, showing that superordinate information in the narrative hierarchical structure is more likely to be recalled than subordinate information (e.g., Horiba, van den Broek, & Fletcher, 1993; van den Broek, Lorch, & Thurlow, 1996).

However, story grammars have a major limitation: The structural analysis of categories does not sufficiently explain what causes a narrative event to occur or what encourages the narrative characters to take subsequent actions. Therefore, the grammar cannot adequately explain how coherence is established through inferences, an important aspect of narrative discourse comprehension.

Causal network analysis

The causal chain approach overcomes the limitation of story grammars. This approach describes a chain of connected narrative events, actions, and states that are directly relevant to the development of the main narrative plot (van den Broek, 1994). Statements in the causal chain (i.e., causal-chain statements) are expected to be represented differently in text memory than statements that are not on the causal chain (i.e., dead-end statements). Some empirical studies have demonstrated that causal-chain statements are included more frequently in summary protocols and receive higher importance ratings than dead-end statements (e.g., Trabasso & van den Broek, 1985).

Causal network analysis (Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985) is a more sophisticated network theory based on the causal chain approach. Network models assume that story events are often derived from multiple casual antecedents, and may have one or more causal consequence. Therefore, each clause in the story is evaluated in terms of its causal relationship to all other clauses, and the network representation of the story is created. Nodes in the network represent the clauses in the story, and arrows capture the causal dependencies between the clauses. The created network representation is evaluated with respect to not only causal chain status but also the number of causal connections of a particular clause to other clauses. Figure 2.1 shows an example of a causal network representation of the story, along with its hierarchical structure. In Figure 2.1, the circled numbers correspond to clauses on the causal chain.

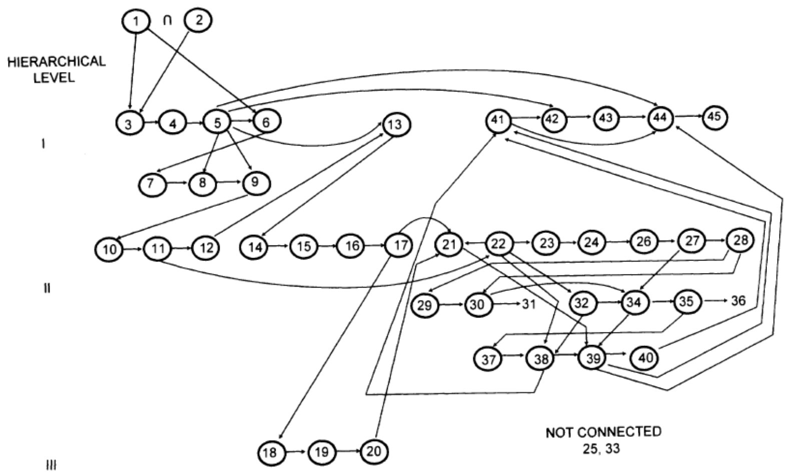

Figure 2.1. Example of a causal network representation of the story (adopted from van den Broek et al., 1996, p. 3014).

Previous studies have found that statements with multiple causal connections are rated as more important, and are therefore more frequently recalled and included in text summaries (e.g., Trabasso & van den Broek,1985; van den Broek, 1988; van den Broek et al., 1996). Van den Broek (1988), for example, found that the importance of a goal statement increased strongly as the number of causal connections that the statement had to other statements increased, independent of its hierarchical level. Van den Broek et al. (1996), for another example, demonstrated that memory of a televised story was strongly influenced by the number of causal connections each event had, but the effects were stronger in adults than children. Trabasso and Wiley (2005, 2009) also demonstrated that narrative structures represented by the causal network can predict the accessibility of text information in readers' text representations. These studies support the validity of using causal network analysis models to capture narrative discourse structure.

Differences between narrative and expository reading

Successful text comprehension means constructing a coherent text representation, regardless of text genre (e.g., narrative, expository). However, previous studies have indicated that some elements differ between narrative and expository reading.

Narrative texts appeals to readers' world knowledge, and are therefore easier to understand and recall than expository texts. In contrast, expository comprehension is often more difficult than narrative comprehension, partly because of the lack of relevant prior knowledge. Indeed, some studies have demonstrated that recall performance among L1 adult readers is much better for narrative than expository texts (e.g., Freedle & Halle, 1979; Graesser, 1981).

In addition, Iseki and Kawasaki (2006) suggested that the situation models constructed from narrative texts have different characteristics to those constructed from expository texts. This study compared how situation models differ between narrative and expository texts in terms of Zwaan, Langston et al.'s (1995) E-I model. They found that the causal dimension contributed to building the situation models of both narrative and expository texts, whereas the intentional dimension contributed only to the construction of narrative situation models. In other words, when reading narrative texts, readers paid most attention to events which were causally related to a protagonist's goal or intention.

Furthermore, Horiba (2000, Experiment 1) indicated that reading narratives and expository texts (e.g., newspaper essays) engage different processing mechanisms. Think-aloud data revealed that L1 readers generated different types of inferences when they read narrative versus expository texts. Specifically, readers generated forward inferences more frequently during narrative than expository reading, whereas backward inferences and general world associations were drawn more frequently during expository than narrative reading.

2.1.3 Differentiating L1 and L2 Reading

The previous sections reviewed several models of reading comprehension and theories of narrative structure analysis developed from L1 research. The basic reading requirements assumed by L1 models and theories can be applied to L2 reading, but L2 reading is presumably a more complicated and demanding process than L1 reading. Although limited in number, the findings from the following studies have yielded important information about the nature of L2 reading (especially L2 discourse processing).

Effects of text structure knowledge on L2 reading

A series of studies by Horiba examined the effects of text structure knowledge on L2 reading. Horiba (1990) compared written recall protocols produced after reading narratives between native speakers of Japanese (L1 readers) and native speakers of English learning Japanese (advanced L2 learners). There was no significant difference in the amount of recalled text statements between the L1 and L2 groups. However, there was a significant difference between groups in the number of causal connections between statements recalled: The L1 group recalled significantly more causal connections than the L2 group. These results suggest that L1 readers constructed more causally coherent text representations than L2 readers.

Horiba et al. (1993) also examined the effects of narrative texts' structural properties (causal factors, story-grammar category, hierarchical level) on recall in L1 and L2 readers. In this study, the L1 group consisted of American college students, while the L2 group consisted of Japanese high school students. Two criteria for scoring recall protocols were used: meaning-preserving recall (i.e., credit was given for a statement only if it was recalled verbatim or in a close paraphrase) and structure-preserving recall (i.e., credit was also given for a statement that served the same structural role as the original text). The results indicated that both L1 and L2 groups performed better on structure-preserving than meaning-preserving recall, suggesting that even L2 readers made use of the text structure to construct a coherent text representation. However, the effects of some structural properties on recall differed between L1 and L2 readers. Both meaning-preserving and structure-preserving recall were influenced by causal connections, story-grammar category, and hierarchical level for L1 readers, whereas recall in L2 readers was only affected by causal connections and story-grammar category when the structure-preserving recall criterion was used. Therefore, the effects of some structural properties, such as hierarchical-level factors, seem to be limited for L2 readers with limited language competence.

Similarly, Horiba (1996) examined recall performance in L2 readers (native speakers of English learning Japanese) after a first and second reading of a narrative text. Although L2 readers were not sensitive to the causal structure of the text during their first reading, proficient L2 readers recalled story events according to the causal structure during the second reading. Horiba noted that the effects of causal structure appeared with delay, presumably because lower level processing (e.g., word recognition, syntactic or semantic analysis of clauses and sentences) was not completed during the first reading.

The results of Horiba's studies suggest that text structure knowledge have some impact, even in L2 reading; however, the effects are smaller or limited compared to L1 reading, presumably because of limited language proficiency in L2 readers.

Effects of text types on L2 reading

Other studies have investigated the effects of text types on L2 reading. Yoshida (2012) examined recall performance after reading narrative and expository texts in Japanese university EFL students. The results showed that recall for the main ideas in each passage was better for narrative versus expository texts both immediately and one week later. This suggests that, as in L1 reading, narrative texts are easier to comprehend than expository texts in L2 reading.

Furthermore, Horiba (2000, Experiment 1) found processing differences based on text type in L2 readers. Participants in this study engaged in a think-aloud task while reading narrative and expository texts. As previously mentioned, the analysis of think-aloud protocols indicated that L1 readers (native Japanese speakers) generated more forward inferences for narrative than expository texts, whereas they generated more backward inferences for expository than narrative texts. In contrast, L2 readers (native English speakers learning Japanese) generated more elaborative inferences when reading expository than narrative texts, thus providing some evidence for text type-appropriate processing among L2 readers. However, it should be noted that the processing differences between narrative and expository texts were rather limited in L2 readers. Regardless of text type, L2 readers used a large portion of their cognitive resources for lower level text processing. This suggests that text processing in L2 readers is influenced by both text type and lower level processing demands.

Strategic processing and updating of situation models during L2 reading

As reviewed above, Horiba's studies (Horiba, 1990, 1996, 2000; Horiba et al., 1993) found that higher level text processing (e.g., the use of structural narrative properties, text type-appropriate processing) is somewhat limited during L2 versus L1 reading due to lower level processing demands in L2 reading. This may be a critical factor that differentiates L2 from L1 reading. Indeed, some previous studies have indicated that some other reading processes, such as strategic reading and updating of situation models, are limited or inhibited during L2 compared to L1 reading due to limited cognitive resources, as described below.

For example, Horiba (2000, Experiment 2) suggested that strategic processing is more limited during L2 compared to L1 reading. In this study, L1 readers (native speakers of Japanese) and L2 readers (native English speakers learning Japanese) were told to read passages either freely (i.e., read normally) or for coherence (i.e., read to pay attention to how the current sentence relates to prior and incoming text). Participants' think-aloud responses indicated that L1 readers in the read-for-coherence condition generated more backward and forward inferences than those in the read-freely condition. In contrast, there were no significant differences for L2 readers in the pattern of text processing between the two reading conditions. Although qualitative analysis suggested that L2 readers tried to alter text processing according to instructions, large amounts of their cognitive resources were allocated to lower level processing, regardless of instruction type. Horiba attributed this result to L2 readers' limited language proficiency, wherein the demands of lower level processing inhibited them from strategically altering their higher level processing mode.

In a more recent study, Horiba (2013) compared the effects of three different task instructions on Japanese university EFL students' reading. Participants were told to (a) pay attention to words and expressions in the text, (b) visualize described events, or (c) compare the author's views with their own. Consistent with Horiba (2000, Experiment 2), analysis of think-aloud data suggested that although readers tried to allocate resources to various levels of processing according as instructed, the effect of task instructions on text processing was limited. This study further revealed that the relationship between task instruction types and these effects on text processing were likely moderated by learners' L2 proficiency and general comprehension skill.

Additionally, Horiba (2013) also suggested that the effects of strategy or task instructions are not clearly observed in reading outcomes (e.g., text comprehension assessed by written recall tasks). Similar results were obtained in Yoshida (2012), where recall performance of Japanese university EFL students was analyzed after engaging in different concurrent reading activities (outlining, answering embedded questions, and reading only).

Ushiro (2010) investigated updating of situation models in L2 readers (Japanese university EFL students). Using passages that were initially misleading about narrative characters, this study examined how L2 readers update their initial interpretation as they process the subsequent passage context. The analysis of several types of data (e.g., sentence reading times and think-aloud protocols) indicated that the extent to which readers successfully revised their text comprehension depended on their L2 reading proficiency. That is, proficient L2 readers updated their text interpretations more flexibly and precisely than less proficient L2 readers.

Similarly, Morishima (2013) examined inconsistency detection in narratives by L2 readers. This study employed an inconsistency detection paradigm (O'Brien et al., 1998), in which the target sentence was either consistent or inconsistent with an earlier portion of text. As mentioned above, prior L1 studies showed that reading times for target sentences are longer when the text contains an inconsistency (e.g., Albrecht & O'Brien, 1993; O'Brien et al., 1998). Morishima replicated this result with L1 readers (native speakers of English), but L2 readers (Japanese university EFL students) failed to detect inconsistencies when even a single sentence separated the inconsistent statements; however, they did detect inconsistencies when the contradicting statements were consecutive. These results suggest that L2 readers have fewer cognitive resources available for global text processing, and consequently have difficulty reactivating earlier text information even when the textual distance is only a couple of sentences.

In addition, Horiba (1990) found that inferences, general knowledge use, and associations were more frequently produced by L1 compared to L2 readers. Similarly, Horiba (1996) found that several types of inferences can be drawn during L2 reading, but the amount of inference generation is limited. These results suggest that L2 readers devote more attention to lower level text processing, and are therefore limited in their ability to generate inferences compared with L1 readers (inference generation in L2 reading is reviewed in detail in 2.2.3).

The findings from previous studies reviewed above suggest that constructing text representations likely differs between L1 and L2 readers. More specifically, higher level text processing, including the use of text structural knowledge, strategic processing, updating situation models, and inference generation, is more limited in L2 than L1 reading due to L2 readers' limited cognitive resources and lower level processing demands. L2 reading research, including this dissertation, should take these previous findings into consideration.

2.2 Inference Generation in Reading Comprehension

2.2.1 Inference Types

As noted in Chapter 1 and the introduction of reading models, inference generation plays a significant role in reading comprehension by establishing textual coherence or embellishing what is explicitly stated in the text. Inference generation in text reading are defined as processes through which (a) readers' relevant background knowledge is activated, and (b) a set of implicit text information is encoded in the text representations (Graesser & Kreuz, 1993; Graesser et al., 1994).

Several previous studies have proposed various classifications of inferences in narrative text comprehension (e.g., Graesser et al., 1994; van den Broek, 1994; van Dijk &

Kintsch, 1983). Van Dijk and Kintsch (1983) classified inferences into two broad categories: *bridging* and *elaborative inferences*. Bridging inferences preserve the local coherence of the passage, and therefore are needed for comprehension. These inferences are considered to be automatically generated during the course of comprehension. Bridging inferences include inferences about anaphor, the causal antecedents of events, and so on. For example, when reading the sentences *The spy threw the report into the fire. The ashes floated up*. (Singer & Ferreira, 1983), readers infer that the report burning is the causal antecedent of the event described in the second sentence. Likewise, when reading the noun phrase *the criminal*, readers infer that it refers to the same person introduced earlier in the story as *the burglar* (McKoon & Ratcliff, 1980).

In contrast to bridging inferences, elaborative inferences are not necessary for comprehension but rather go beyond or embellish what has been explicitly stated. Therefore, unlike bridging inferences, elaborative inferences are not automatically generated during comprehension. Inferences such as those about the causal consequences of events and the emotions of a protagonist are included in elaborative inferences. For example, when reading the sentence *The director and cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story* (McKoon & Ratcliff, 1986), readers would predict the actress's death as a consequence of the event described in the sentence. Similarly, when reading the sentence *Suzan had trouble retaining her tears when thinking of her grandmother alone in her room*, readers infer that Suzan was feeling sad (Gygax, Tapiero, & Carruzzo, 2007).

Van den Broek (1994) offered three major inferential processes during text comprehension in terms of the direction in which inferences are generated: *forward*, *orthogonal*, and *backward* elaborations (see Figure 2.2). Forward elaborations concern the anticipation of future events and expectations about the future importance of focal events. Orthogonal elaborations embellish the focal events by adding details or associations (e.g., spatial and visual information). Backward elaborations, which are the most important for comprehension, connect the focal statement to events or states previously described in the text.

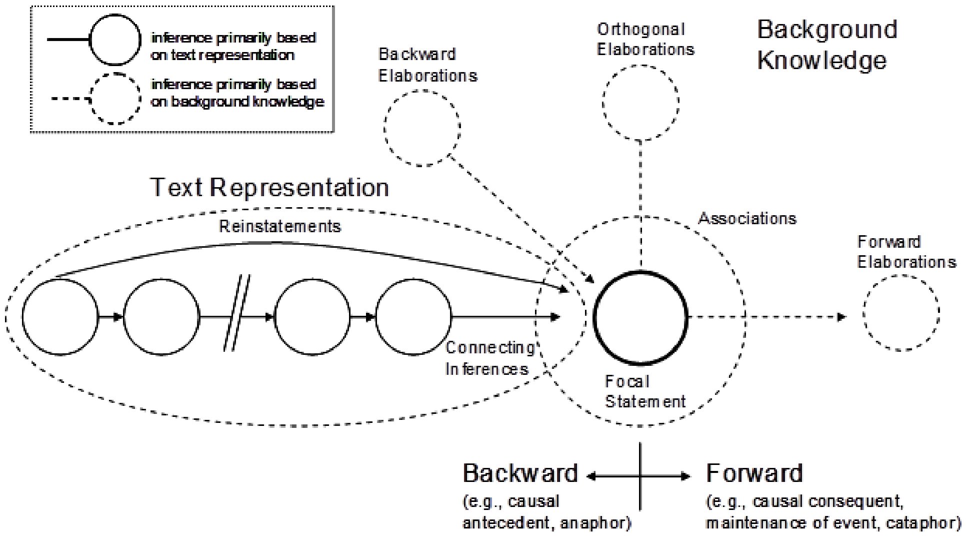

Figure 2.2. A process model of inference generation during reading (adopted from van den Broek, 1994, p. 577).

More detailed classification has been suggested by Graesser et al. (1994). They proposed 13 classes of inferences as listed in Table 2.1. This classification is "defined according to the contents of the inference and its relation to the explicit text" (Graesser et al., 1994, p. 375). Based on their explanation, Classes 1 (referential), 2 (case structure role assignment), and 3 (causal antecedent) are necessary for maintaining the local coherence of a text, whereas Classes 4 (superordinate goal), 5 (thematic), and 6 (character emotional reaction) establish global coherence. In addition, Classes 3 and 4 provide explanations of causes of events or a character's actions described in the text. Classes 7 to 11 (causal consequence, instantiation of noun category, instrument, subordinate goal, and state) are elaborative inferences, and Classes 12 (emotion of reader) and 13 (author's intent) are about the communication between reader and author.

Table 2.1

Thirteen Classes of Inferences (Adapted From Graesser et al., 1994, p. 375)

Type of inferences		Brief description
Class 1:	Referential	A word or phrases referentially tied to a previous element or constituent in the text (explicit or inferred).
Class 2:	Case structure role assignment	An explicit noun phrase is assigned to a particular case structure role (e.g., agent, recipient, object, location, time).
Class 3:	Causal antecedent	The inference is on a causal chain (bridge) between the current explicit action, event, or state and the previous passage context.
Class 4:	Superordinate goal	The inference is a goal that motivates an agent's intentional action.
Class 5:	Thematic	This is a main point or moral of the text.
Class 6:	Character emotional reaction	The inference is an emotion experienced by a character, caused by or in response to an event or action.
Class 7:	Causal consequence	The inference is on a forecasted causal chain, including physical events and new plans of agents. These inferences do not include the character emotions in class 6.
Class 8:	Instantiation of noun category	The inference is a subcategory or a particular exemplar that instantiates an explicit noun or an implicit case role that is required by the verb.
Class 9:	Instrument	The inference is an object, pan of the body, or resource used when an agent executes an intentional action.
Class 10:	Subordinate goal-action	The inference is a goal, plan, or action that specifies how an agent's action is achieved.
Class 11:	State	The inference is an ongoing state, from the time frame of the text, that is cot causally related to the story plot The states include an agent's traits, knowledge, and beliefs; the properties of objects and concepts; and the spatial location of entities.
Class 12:	Emotion of reader	The inference is the emotion that the reader experiences when reading a text
Class 13:	Author's intent	The inference is the author's attitude or motive in writing.

2.2.2 Theories of Inference Generation

Because text processing is constrained by limited cognitive resources, it is generally assumed that only a subset of all possible inferences is automatically generated during reading. Therefore, a central issue for reading researchers is what types of inferences are automatically generated during the course of reading (i.e., *on-line inference generation*). There have been two major theoretical positions that predict on-line inference generation during text processing: the *minimalist hypothesis* (McKoon & Ratcliff, 1992) and the *constructionist theory* (Graesser et al., 1994).

According to the minimalist hypothesis, only two types of inferences are drawn on-line: (a) inferences based on easily available information from general knowledge and explicit information from the text, and (b) inferences needed to maintain the local coherence of the text. Thus, the minimalist hypothesis proposes that the number of inferences generated during reading is limited. From the viewpoint of Graesser et al.'s (1994) categorization, Classes 1 to 3 (inferences about the referential word or phrase, case structure role, and causal antecedents) are expected to be generated on-line (see Table 2.2). McKoon and Ratcliff (1992) provided evidence for this hypothesis in a series of experiments. Although the minimalist hypothesis does not specify what types of inferences are generated strategically during reading, they suggested that strategic inferences (a) need more time to be constructed, (b) are drawn when local coherence cannot be established, and (c) are not consistently drawn and are influenced by reader characteristics, goals, and tasks.

The constructionist theory suggests which inferences are generated on-line based on three critical assumptions: (a) the reader goal assumption (i.e., the reader builds a representation that addresses their goals), (b) the coherence assumption (i.e., the reader tries to develop a representation which is both locally and globally coherent), and (c) the explanation assumption (i.e., the reader attempts to explain why a character's actions or events are mentioned in the text). Therefore, in contrast to the minimalist hypothesis, the constructionist theory predicts that Classes 1 to 6 are automatically generated during reading (see Table 2.2 again). The predictions of this theory have been confirmed empirically (e.g., Long & Golding, 1993; Long, Golding, & Graesser, 1992; Sue & Trabasso, 1993). Although the constructionist theory predicts that other inference classes will not be generated on-line, it suggests that elaborative inferences (Classes 7 through 11) can be generated on-line only under two conditions: (a) when the reader has a specific goal in generating the inferences, and (b) when the possible inferences are strongly constrained by context.

Table 2.2

Predictions of On-Line Inference Processing, Derived From Minimalist Hypothesis and Constructionist Theory (Adapted From Graesser et al., 1994, p. 384)

		Minimalist hypothesis	Constructionist theory
Class 1:	Referential	X	X
Class 2:	Case structure role assignment	X	X
Class 3:	Causal antecedent	X	X
Class 4:	Superordinate goal		X
Class 5:	Thematic		X
Class 6:	Character emotional reaction		X
Class 7:	Causal consequence		
Class 8:	Instantiation of noun category		
Class 9:	Instrument		
Class 10:	Subordinate goal-action		
Class 11:	State		
Class 12:	Reader emotion		
Class 13:	Author intent		

Note. X = on-line prediction.

2.2.3 Inference Generation in L2 Reading

Inference generation is also of critical importance in L2 reading, both theoretically and educationally. Koda (2005) asserted that future L2 studies should address “the specific process involved in inferential generation among L2 readers or the conditions affecting it" (p. 130). Additionally, Chikalanga (1992) suggested that L2 teachers should know what types of inferences can be produced in reading. To date, some studies have investigated inference generation in the L2 reading context, though the number of studies is limited. These studies investigated whether L2 readers generate inferences, and what types of inferences are generated during L2 reading.

For instance, Horiba's (1996) think-aloud data indicated that L2 readers (native English speakers leaning Japanese) make several types of inferences during reading, such as backward and forward inferences. However, this study revealed the effects of the level of readers' L2 proficiency on their inference generation: L2-advanced readers generated backward and forward inferences as frequently as L1 readers, whereas L2-intermediate readers did not generate these inferences. This study attributed the limited production of inferences among L2-intermediate readers to the fact that these readers needed to allocate more cognitive resources to lower level reading processes such as word recognition and syntactic analysis. As a result, L2-intermediate readers did not have sufficient cognitive resources to engage in higher level processing such as inference generation.

Yoshida (2003) found similar effects of L2 proficiency on inference types generated during reading. In this study, Japanese university EFL students read the same passage as in Horiba (1996) and engaged in the think-aloud task. The think-aloud data showed that readers with higher L2 proficiency more frequently generated elaborative inferences (e.g., superordinate goal, causal consequence) than readers with lower L2 proficiency, whereas the number of bridging inferences (e.g., causal antecedent, subordinate goal) was not related to readers' L2 proficiency. Yoshida explained that proficient L2 readers made more elaborative inferences because generating these inferences required them to sufficiently comprehend the text meaning.

Yoshida (2003) also indicated the effects of readers' working memory (WM) capacity on the number of inferences generated during reading. The results demonstrated that readers with high WM capacity generated a larger number of inferences than those with low WM capacity, regardless of the inference types. Yoshida suggested that readers with low WM capacity had insufficient cognitive resources for higher level reading processing, that is, inference generation.

Muramoto (2000) investigated the generation of four types of inferences (i.e., goal, action, emotion, and state) in L2 reading. In this study, Japanese university EFL students read some narrative texts, and then some target sentences were presented to the participants. They were required to judge whether the sentence had appeared in the text they had just read (i.e., a sentence recognition task; see also 2.3.2.2). The results of the sentence recognition task demonstrated that readers with higher L2 proficiency are more likely to make inferences than lower L2 proficiency readers, in line with Horiba's (1996) findings. This trend was consistent between four types of inferences.

These studies indicate that L2 readers make some inferences during reading and L2 proficiency plays a role in inference generation. However, the aforementioned studies investigated the generation of several types of inferences produced while reading a passage, but have not closely examined the specific types of inference drawn during L2 reading. In contrast, Collins and Tajika (1996) examined whether Japanese university EFL students make instrumental inferences during reading, and provided some evidence for inference generation using memory tests (i.e., word stem completion and cued recall). In more recent research, Shimizu (2012) focused on bridging inference generation in L2 reading. A series of experiments showed that Japanese EFL learners generated bridging inferences during narrative reading, and L2 proficiency influenced the likelihood of bridging inference generation.

Nevertheless, many researchers in the L1 reading area have focused on *predictive inferences* (also called "forward elaborations" in van den Broek, 1994, and "causal consequence inferences" in Graesser et al., 1994), which has not closely investigated in the L2 context. The following section provides the detailed review of previous studies on predictive inference generation during reading.

2.3 Predictive Inferences in Reading Comprehension

2.3.1 Role of Predictive Inferences

Predictive inference is a typical type of elaborative inference and concerns the likely outcome of events or actions described in the preceding passage. For instance, when reading the sentence *No longer able to control his anger, he threw a delicate porcelain vase against the wall* (Klin, Guzmán, & Levine, 1999), readers would predict that the vase will break (i.e., activate a predictive inference). Linderholm (2002) claimed that predictive inferences play an important role in reading comprehension for at least three reasons:

> First, to make predictive inferences, readers must connect text events with background knowledge and this allows readers to anticipate what is likely to occur next. The mere connection of explicit text events with background knowledge is a key step toward building situation models of texts, which is the ideal form of mental representation for long-term retention. Second, the generation of predictive inferences eases the processing of future text events. For example, if a reader anticipates an upcoming event, once that event is confirmed, reading proceeds smoothly and information about the predicted event becomes easily integrated into the reader's situation model. Third, there are some classes of predictive inferences that must be made to maintain text coherence during reading. Thus, predictive inferences can be seen as an important process for

forming situation models of text information, processing future text more smoothly, and maintaining coherence of text events during reading (p. 260).

In line with the idea suggested in Linderholm (2002), Fincher-Kiefer (1993) demonstrated that the development of situation models involves elaborative inferential processing, including predictive inference generation. In addition, it has also been claimed that making predictive inferences during reading represents a form of active engagement with the text (Allbritton, 2004) and increases readers' cognitive interest (Campion, Martins, & Wilhelm, 2009). Thus, many researchers agree that readers can benefit from making predictive inferences during reading. Indeed, from the pedagogical viewpoint, Day and Park (2005) suggested that asking students what might happen next in a described situation helps them become more interactive readers. Similarly, making predictions while reading is often regarded as an effective comprehension strategy (e.g., Carrell, 1989; Palincsar & Brown, 1984).

2.3.2 Making Predictive Inferences

A large number of studies have discussed whether L1 readers make predictive inferences on-line. Some researchers have provided evidence that these inferences are drawn during L1 reading (e.g., Fincher-Kiefer, 1995, 1996), whereas others have not (e.g., Magliano, Baggett, Johnson, & Graesser, 1993; Singer 1980).

As noted in 2.2.2, both the minimalist hypothesis (McKoon & Ratcliff, 1992) and the constructionist theory (Graesser et al., 1994) predict that these inferences are unlikely to be made during reading, but are probably made under some other conditions. In line with this prediction, more recent research has identified the several factors that affect the generation of predictive inferences, such as (a) contextual constraint (Cook, Limber, & O'Brien, 2001; Klin, Guzmán, et al., 1999; Linderholm, 2002), (b) whether the inferences are related to narrative characters' motivation or goals (Klin, Murray, et al., 1999; Murray, Klin, & Myers, 1993), (c) availability of textual information (Keefe & McDaniel, 1993; Murray et al., 1993), (d) reading skills (Murray & Burke, 2003), (e) WM capacity (Linderholm, 2002; Virtue, van den Broek, & Linderholm, 2006), (f) time required to construct inferences (Calvo & Castillo, 1996, 1998; Calvo, Castillo, & Estevez, 1999), and (g) the reading goal or strategy (Allbritton, 2004; Magliano, Trabasso, & Graesser, 1999).

In addition, some researchers have claimed that methodological variation between previous studies affected the results (Keenan, Potts, Golding, & Jennings, 1990; Potts, Keenan, & Golding, 1988). The following section describes several factors that affect predictive inference generation during reading as well as some measurements used to detect these inferences.

2.3.2.1 Factors Affecting Predictive Inference Generation

Contextual constraint

Many previous studies have focused on the effects of text factors on predictive inference generation. Among various text factors, the effects of contextual constraint (i.e., the degree to which the text constrains a possible inference) have been extensively investigated in past L1 studies (e.g., Cook et al., 2001; Lassonde & O'Brien, 2009; Linderholm, 2002; Virtue et al., 2006). Previous studies have demonstrated that predictive inferences are more likely to be drawn during reading when the context strongly constrains the possible inference than when it weakly constrains the inference. Let us consider the example of throwing a vase, as described above. If reading about a plastic vase being thrown against the wall instead of a delicate porcelain vase (i.e., a weak contextual constraint condition), readers would equally predict two consequences: The vase will or will not break. Thus, in the case of weak contextual constraint, the activation from the predictive sentence is assumed to be divided between these two consequences, resulting in lower likelihood of inference generation compared to the case of strong contextual constraint (Casteel, 2007; Klin, Guzmán, et al., 1999). Therefore, contextual constraint of the passage can be a significant variable for determining the likelihood of predictive inference generation.

Relation to the narrative characters' goals or motivation

Several studies also have suggested that predictive inferences are more likely to be generated when these inferences are related to narrative characters' goals or motivation. Klin and her colleagues (Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999) suggested two subtypes of predictive inferences, depending on whether or not the inferences are related to characters' goals or motivation. These subtypes are *consequence* and *motivational* predictive inferences. The consequence inference is purely the result of the events described in the passage. The first passage in Table 2.3 shows an example for consequence inferences. The expected inference for this passage (i.e., the vase breaking) is the highly probable consequence of the vase being thrown against the wall, but is less related to the character's goals or motivation. This kind of predictive inference only serves to embellish the passage; it is not necessary to preserve the local coherence of the text.

In contrast, the motivational inference is not only a possible outcome of the described event, but also the motivation or cause of the described character's action. The second passage in Table 2.3 shows an example for motivational inferences. The expected inference in this passage (i.e., stealing the ring) is not only the likely consequence of Brad quietly making his way to the counter, but also the motivation for his doing so. In other words, the inference explains why the character took the intentional action described in the passage. Thus, this inference is necessary to maintain the local coherence of the text.[1]

Table 2.3

Two Subtypes of Predictive Inferences (Adapted From Klin, Murray, et al., 1999, p. 244)

1. Consequence predictive inference
Steven had been married for years, and his resentment had been building up. One day, no longer able to control his anger, he threw a delicate porcelain vase against the wall.
Inference: *The vase will break.*
2. Motivational predictive inference
Brad had no money but he just had to have the beautiful ruby ring for his wife. Seeing no salespeople around, he quietly made his way closer to the counter.
Inference: *He will steal the ring.*

Klin and colleagues (Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999) found that consequence inferences are only drawn when they are strongly constrained by the context, whereas motivational inferences are made regardless of the strength of contextual constraint. Thus, evidence has shown that readers are more likely to generate motivational inferences than consequence inferences.

Availability of textual information

Some research has also indicated that the availability of textual information is important for constructing inferences. For example, Keefe and McDaniel (1993, Experiment

[1] Some researchers may disagree with the inclusion of motivational inferences within the class of predictive inferences because predictive inferences are traditionally considered unnecessary to maintain the local coherence of the text (e.g., Graesser et al., 1994). However, other researchers have investigated motivational inferences as a subtype of predictive inferences because they represent the prediction of a likely consequence of the described event (Allbritton, 2004; Klin, Murray et al., 1999; Murray & Burke, 2003).

3) examined the effects of information availability on predictive inference generation. One group of participants performed a naming task for inference-related words immediately after the sentence inducing predictive inferences (e.g., *After standing through the three-hour debate, the tired speaker walked over to his chair*; the inference-related word was *sat*.). Another group of participants performed the same task after an intervening sentence that followed the predictive sentence (e.g., *He realized that his valiant effort was probably in vain*). An analysis of naming latencies suggested that predictive inferences were activated immediately when the predictive sentence was read, whereas the inferences were not detected when the intervening sentence followed the predictive sentence. No evidence for inferences was obtained in a third group of participants who counted backwards in the time period between the predictive sentence and inference-related word instead of reading an intervening sentence. These results suggest that reading an intervening sentence or counting backwards reduced the availability of the predictive sentence in WM, preventing predictive inference generation.

Murray et al. (1993) also emphasized that the availability of relevant text information is important for constructing predictive inferences. These researchers noted that evidence for predictive inferences was not obtained in some previous research (e.g., Potts et al., 1988) because the experimental materials did not allow information relevant to inferences to be maintained. Indeed, Murray et al. found evidence for predictive inferences when materials used in previous research were rewritten so that the relevant text information could be maintained and was available during reading. Although this study also examined the effects of passage length and coherence break on predictive inference generation, it is concluded that these factors are not critical compared to the availability of information relevant to inferences.

Reading skills and WM capacity

Furthermore, researchers have identified some reader variables related to the activation of predictive inferences. For instance, Murray and Burke (2003) demonstrated that skilled readers showed shorter naming latencies for inference-related words for the predictive context than control context, but less-skilled readers did not. This suggests that reading ability affects the likelihood of activation of predictive inferences.

Linderholm (2002) found that readers with high WM capacity were more likely to activate predictive inferences than those with low WM capacity. Linderholm suggested that readers with high WM capacity are more likely to possess sufficient information for predictive inference generation than readers with low WM capacity because they have more cognitive resources available during reading.

In relation to WM capacity, some studies demonstrated that the amount of cognitive resources available during reading affect inference generation. For instance, Shears et al. (2007) examined the effects of cognitive demand on bridging inference generation during reading by employing dual-task methodology, in which participants are required to hold a list of words in memory while reading some passages. The analysis of recognition times for the target word and knowledge-validating questions demonstrated that there was some impairment in inference activation under the dual-task condition compared to the single task condition. This suggests that readers allocated their cognitive resources to holding the word lists in memory and few resources were available for inference generation.

Fincher-Kiefer and D'Agostino (2004) used a dual-task methodology to examine predictive inference processing when cognitive resources were divided between two tasks. They found evidence for predictive inference activation when readers were given a verbal memory load (i.e., holding a string of letters in memory), but not when they were given a perceptual memory load (i.e., holding a dot array in memory). These results suggest that making predictive inferences during reading requires visuospatial resources in particular.

Time for constructing inferences

Research has shown that predictive inferences take time to construct and occur with delay. Calvo and Castillo (1996, 1998) and Calvo et al. (1999) used word-by-word, fixed-pace, rapid serial visual presentation (RSVP) of the context sentence, and manipulated stimulus onset asynchrony (SOA; i.e., the interval between the onset of the last context word and the onset of the target word). Facilitation in naming times for the target word was observed for SOAs of 1,250 and 1,500 ms, but not 500 or 1,000 ms. This suggests that making predictive inferences requires about one sec after the predictive context.

Other studies have used self-paced reading or presented passages one sentence or line at a time (Campion & Rossi, 2001; Cook et al., 2001; Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999; Virtue et al., 2006). These studies found that predictive inferences were detected after a short blank inter-stimulus interval (ISI; 250–1,500 ms) between the predictive context and the target word. However, Calvo, Castillo, and Schmalhofer (2006) noted that it is difficult to determine the time course of predictive inferences with this procedure because participants can decide when to finish reading the predictive sentence before the inference word appears.

In an alternative self-paced reading procedure, Calvo and Castillo (1998, Experiment 2) collected reading times during the presentation of predictive or control sentences (e.g., *Three days before the examination the student went to the library, looked for a separate table and opened his notebook*; *The student, who was very tired after finishing his examination, forgot his notebook and left it in the library*) and a continuation sentence containing the inferential word (e.g., *The student <u>studied</u> for an hour approximately*; the inferential word is underlined). Using a moving-window technique, they only observed facilitation for reading the post-target and final regions of the continuation sentence, but not the target word itself, in predictive compared to control contexts. Furthermore, based on assessments of eye movements, Calvo, Meseguer, and Carreiras (2001) found that facilitation for target words was observed in late processing, not early processing (see also 2.3.2.2 for the eye tracking method). Thus, the results of both studies suggest that constructing predictive inferences takes time.

Some studies also have suggested effects of reader variables on the time course of predictive inference generation. For example, Linderholm (2002) suggested that low-WM capacity readers need more time to generate predictive inferences than high-WM capacity readers. Similarly, Murray and Burke (2003) noted that less skilled readers may require more time to construct inferences than skilled readers.

Reading goals or strategies

Other studies have investigated the effects of reading goals or strategy instructions on predictive inference generation (e.g., Allbritton, 2004; Magliano et al., 1999; van den Broek, Lorch, Linderholm, & Gustafson, 2001). These studies often involve a pre-reading instructional set to encourage strategic processing. For example, in Magliano et al. (1999), participants were instructed to read passages for explanation, prediction, association, or understanding. The analysis of think-aloud protocols demonstrated that participants strategically controlled the inferences they generated according to instructions. Specifically, participants produced more predictive inferences in the prediction condition than in the understanding condition. In addition, van den Broek et al.'s (2001) analysis of think-aloud protocols revealed that participants made more predictive inferences when reading passages for study versus entertainment. Furthermore, Allbritton (2004) investigated the strategic production of predictive inferences using a lexical decision task and manipulating the post-reading task. In this study, readers provided evidence of predictive inference generation when they were asked to write a sentence that continued the story, but not when asked to answer a literal comprehension question about the story. In short, previous studies suggested that predictive inference generation during reading can be facilitated by giving reading goals or tasks encouraging readers to strategically process the text for the inference generation.

Additionally, some previous studies examined the effects of strategic processing aimed at predictive inferences on comprehension of explicit text comprehension. They found that strategic processing does not affect explicit text comprehension, as assessed by written recall (Magliano et al., 1999) or yes/no comprehension tests (Calvo et al., 2006).

However, it should be noted again that effects of strategic processing on inference generation were limited in L2 reading (see 2.1.3). The analysis of think-aloud protocols in Horiba (2000, Experiment 2) found no significant differences in the pattern of lower and higher level processing between the two reading task conditions (i.e., read freely and read for coherence). Horiba suggested that limited cognitive resources among L2 readers inhibited them from strategically altering their higher level processing mode, even though they were aware of the instructions.

2.3.2.2 Measures of Inference Generation

A number of measures have been used to assess inference generation during reading. Measurers of inference generation can be divided into two major classes based on testing time: *off-line* and *on-line*. Off-line measures assess inferences during reading after a delay, and assess whether the inferential information was encoded as part of long-term text memory (i.e., the complete situation model). These measures include cued recall, sentence recognition, and probability judgment tasks. In contrast, on-line measures test inferences either while or immediately after reading a word or sentence. They include think-aloud protocols, word or sentence reading times, recognition times, naming or lexical decision times, meaningfulness judgment times, and eye fixation durations. These measures are reviewed below in the context of prior studies that primarily focused on predictive inferences.

Cued recall

The most frequently used off-line measure is cued recall. In this task, participants read sentences with implicit or explicit information (e.g., *The director and cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story* or *…the actress fell from the 14th story* <u>*and died*</u>; McKoon & Ratcliff, 1986). After reading a set of such sentences, the inference words are then presented as retrieval cues to recall the sentences. Several studies have shown that recall performance does not differ, regardless of whether the target information was implicitly suggested or explicitly mentioned in the text. This suggests that readers made inferences represented by the target words and encoded these inferences into text memory in the same manner as explicit information (e.g., McKoon & Ratcliff, 1986; Paris & Lindauer, 1976; Suzuki, 2000).

Some studies on predictive inferences have used another type of cued recall task (e.g., Klin, Murray, et al., 1999; Murray & Burke, 2003). In these studies, after reading several passages that induce predictive inferences, the first sentences of each passage are presented as retrieval cues to recall the subsequent sentences. If the target inference becomes part of text memory, participants should produce predictive information at the end of recall (e.g., *….the actress fell from the 14th story* <u>*and died*</u>), even though this information was not explicitly described in the text. Thus, recall protocols are scored in terms of whether the target inference is included.

Sentence recognition

The sentence recognition task (SRT) also tests whether inferences were encoded into long-term text memory. Muramoto (2000) presented three types of target sentences to participants after they read several passages: (a) an *explicit* sentence, which described an event explicitly mentioned in the text; (b) an *inference* sentence, which described an event that can be inferred from the text; and (c) an *inconsistent* sentence, which described an event not mentioned or suggested in the text. All of the target sentences were written in participants' L1 to ensure that participants' surface text memory about word forms and sentence structures did not have an effect on recognition judgment. The participants were required to judge whether each target sentence had appeared in the texts they had just read. They were also asked to make confidence ratings for their recognition judgments on a 4-point scale (see Table 2.4). Collecting confidence ratings allows for a finer-grained analysis of recognition data. In this task, recognition scores are calculated from participants' responses to the target sentences and their confidence rating (see Table 2.4). These scores range from 0 to 6. When participants recognize the target sentence as written in the text, the scores are 3–6 according to their confidence rating. In contrast, when participants recognize a target sentence as not written in the text, the scores are 0–3 depending on their confidence rating. If participants

Table 2.4

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

Response	Confidence rating		Recognition scores
This sentence was <u>written</u> in the text.	4	(high)	6
	3	(relatively high)	5
	2	(relatively low)	4
	1	(low)	3
This sentence was <u>not written</u> in the text.	1	(low)	3
	2	(relatively low)	2
	3	(relatively high)	1
	4	(high)	0

make the target inference and encode it as part of their text memory, inference sentences (e.g., *The actress died*) should be falsely recognized, possibly with high confidence. Therefore, higher recognition scores can be taken as evidence for inference generation.

Probability judgments

Campion and Rossi (2001) used the probability judgment task to evaluate predictive inference encoding after reading. After participants read experimental passages, several target sentences are presented. These sentences express predictable events for each passage (e.g., *The actress died*). Participants have to decide if the judgment sentence reflects an event that could probably happen next in the situation described in one of the texts they read. When participants answer "yes," they are required to rate the probability of the event on a 5-point scale (1 = *minimal probability*, 5 = *maximum probability*). If participants integrate target inferences into long-term memory, they should respond "yes" to judgment sentences; moreover, the sentence should be evaluated as highly plausible. This task has also been frequently used to validate predictive inferences in experimental materials in preliminary studies (e.g., Casteel, 2007; Cook et al., 2001; Klin, Murray, et al., 1999; Linderholm, 2002).

These off-line measures are considered to be sensitive to integration or encoding of target inferences into long-term text memory. However, these measures share a limitation: Off-line measures do not discriminate between inferences drawn during reading and those drawn during testing. That is, even if these measures detect an inference, it is possible that the inference was reconstructed at test (e.g., during the recall, recognition, and/or judgment tests) rather than during reading. To ensure that the results of off-line tests reflect processes that occur during reading, researchers have used on-line measures as well as, or instead of, off-line measures. Several on-line measures are reviewed below.

Think-aloud

Think-aloud methodology has been frequently used to investigate on-line comprehension processes in reading (e.g., Horiba, 1996, 2000; Magliano et al., 1999; van den Broek et al., 2001). Participants are asked to verbalize whatever they are thinking while reading a passage. Participants' verbal responses are recorded and divided into separate statements. Then, statements are classified into categories, such as word analysis, sentence analysis, inferences, and self-monitoring, for later analysis (Horiba, 1996, 2000).

Although think-aloud methodology has been accepted as a valid tool for detecting inferences that occur during reading, there are some limitations. One limitation is that thinking aloud reveals little information about passive reading comprehension processes. In other words, the content of verbal reports only reflects what is consciously accessible and verbalizable. Another limitation is that thinking aloud may require readers to adopt a processing mode that differs from normal reading to verbalize their thoughts. This might cause inferences to be drawn that are not normally generated while reading. Therefore, it is essential to validate think-aloud data with other measures, such as written recall or reading times (Horiba, 1996, 2000, 2013; Magliano et al., 1999).

Word or sentence reading times

The word or sentence reading paradigm has been used in several studies to detect on-line inferences. For instance, in Calvo and Castillo (1996, 1998), participants read the predictive context, followed by sentences where a target word either confirmed or disconfirmed the inference suggested by the predictive context (see also *Time for constructing inferences* in 2.3.2.1). Participants read sentences one to four words at a time, and reading times for the target word, the post-target region, and the final region of the continuing sentence were collected. The analysis of reading times showed that reading was facilitated for the post-target and final regions of the continuing sentence when the predicted event was confirmed, and reading was inhibited when the event was disconfirmed.

Other studies have used sentence reading times as an indicator of on-line predictive inference generation (Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999; Murray & Burke, 2003). These studies employed an inconsistency detection paradigm similar to O'Brien et al. (1998). Participants read a passage sentence-by-sentence in which the predictive context (e.g., *Seeing no salespeople or customers around, he quietly made his way closer to the counter*) was immediately followed by a sentence that contradicted the inference induced by the predictive context (e.g., *Staring intensely at the ring, he promised to buy it one day*). Reading times should increase for the contradictory sentence if readers draw the target inference and maintain it in WM. Klin and colleagues (Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999) found evidence for predictive inference generation using this paradigm.

The advantage of the reading time paradigm is that it detects inferences while participants are reading the passage. Consequently, unlike off-line measures, it rules out the possibility that the detected inferences were made during reconstruction (e.g., at the time of recall). However, a disadvantage of this paradigm is that target sentence reading times may be affected by syntax complexity or sentence compatibility in addition to inference generation. Another disadvantage is that reading times do not reveal what inference was drawn. In other words, it is difficult to use reading times to reveal the specific inference detected. In this regard, the remaining measures described below provide more insight into the content of the inference.

Recognition times

In the recognition task, a target word is presented immediately after participants read the predictive or control context. Participants are then required to judge whether the target word appeared in the text they just read as quickly and accurately as possible. The target word represents the specific inference concept (e.g., *dead*) suggested by the predictive context, but does not actually appear in the text.

The logic of this paradigm is that if an inference was generated, the concept will be activated and incorporated into the representation of the predictive context, but not the control or neutral context. Consequently, a correct response to the target word (i.e., "no") would be more difficult (i.e., less accurate or slower) after reading the inference versus control context (McKoon & Ratcliff, 1986; Shears et al., 2007) because readers have some difficulty distinguishing inferential information from information explicitly described in the text at the time of recognition (Fincher-Kiefer, 1995, 1996).

Recognition is sometimes classified as an off-line measure because it is a memory test. Although the terms *on-line* and *off-line* are not always well-distinguished, Jiang (2012) defined *on-line* as follows:

> The term *on-line* is traditionally used to refer to the examination of sentence or discourse comprehension while the comprehension is on-going […] rather than after its completion. This usually means that a probe or data collection has to occur before the end of a sentence or discourse. However, the term on-line has also been used more liberally to refer to task that require fast or temporally constrained responses and measure RTs as data. […] All tasks that require fast responses and produce response time data can be considered on-line tasks. […] Some tasks are more on-line than others, but these RT tasks share the common feature that responses are observed in close temporal proximity to the mental processes under examination. (p. 5)

According to the above definition of the term *on-line*, the current dissertation includes recognition as an on-line measure when it is performed immediately after reading the context, it requires quick judgment, and reaction times are analyzed.

Recognition is sometimes preferable to sentence reading times or think-aloud methodology because the target word can represent a particular inference concept, and the task does not encourage participants to process the text strategically for inferences. However, the major problem with recognition tasks is that they require participants to check the target word against the text (i.e., *context checking*). Consequently, it is difficult to eliminate the possibility that the detected inferences are drawn during the recognition test rather than during reading.

Lexical decision and naming times

One way to confirm that detected inferences are drawn during reading is to use a task that does not ask participants to evaluate the target word against the text, such as lexical decision and naming tasks. These tasks are the most frequently used on-line tools for measuring predictive inference generation in reading. (e.g., Allbritton, 2004; Calvo et al., 2006; Campion & Rossi, 2001; Cook et al., 2001; Virtue et al., 2006). In lexical decision or naming tests, participants read either an inference or control version of a text. They then decide whether a presented target word is a real word (lexical decision) or they are asked to say it out loud (naming) as accurately and quickly as possible. Thus, in these tasks, participants have no apparent reason to compare target probes to the texts. If the inference was activated, lexical access will be primed, facilitating correct responses to the target word. Therefore, correct lexical decision or naming times should be faster for inference than control or neutral contexts.

In order to compare the strength of the inference activation among conditions, *inference activation scores* (Graesser, Wiemer-Hastings, & Wiemer-Hastings, 2001; Virtue et al., 2006) are often calculated from lexical decision or naming latencies. These scores are first calculated for individuals by subtracting a participant’s mean correct response time for inference texts from the mean time for control or neutral texts. For example, if a participant’s mean response time was 700 ms for the inference texts and 900 ms for the control texts, the mean inference activation score for the participant would be 200 ms. Overall, mean activation scores are then calculated for each condition. The score would be significantly greater than zero if inferences are activated. Additionally, the more strongly the inferences are activated, the greater the score would be. Calculating these scores allowed for a direct and clear comparison of inference activation strength between conditions.

Some L2 studies have used the lexical decision to investigate learners’ lexical knowledge or incidental vocabulary learning (e.g., Elgort, 2011; Sonbul & Schmitt, 2013). However, obtaining reliable priming effects of L2 lexical decisions requires a minimum level of L2 proficiency among participants because lexical decisions “rely on the participants’ ability to access and process lexical representations in an automatic manner, with a reasonable degree of accuracy” (Elgort, 2011, p. 371). Studies also have identified word-related factors that affect L2 lexical decisions, such as word length, frequency, and familiarity (de Groot, Borgwaldt, & van den Eijnden, 2002; Yokokawa, 2006). Therefore, these factors as well as participants’ proficiency should be carefully considered when lexical decision is used in the L2 context.

Meaningfulness judgment times

Iseki (2006) used the meaningfulness judgment task (MJT) to test on-line inference generation. As in lexical decision or naming tests, participants read either an inference or control version of a text before performing the task. In the MJT, however, the target probe is a single sentence rather than a single word; the target sentence describes the event that can be inferred from the text (e.g., *The vase was broken*). The MJT requires participants to decide whether a target sentence makes sense or is semantically acceptable. For example, *The vase was broken* is an acceptable sentence; *The vase was cried* is an unacceptable sentence. If the inference has been activated, then the correct response to the target sentence (i.e., yes) should be facilitated. Meaningfulness judgment (also called *acceptability judgment*) is similar to grammaticality judgment; however, it is a broader concept (Jiang, 2012).

Importantly, Iseki (2006) suggested that the task is more sensitive to the degree of inference activation. When the presented probe is a sentence, there is a greater degree of overlap between a probe and an inferential concept than when the probe is a word. Consequently, priming effects for target probes may be larger in the MJT than in other tasks, such as lexical decision or naming tasks.

Eye tracking

Although the aforementioned on-line measures have been frequently used, the tasks employed do not allow participants to read passages in the same way that they do in normal circumstances. For example, in these tasks reading is sentence-by-sentence or word by word, and this requires participants to engage in additional tasks, such as pressing a button frequently. To overcome these shortcomings, researchers have used eye tracking (e.g., Kaakinen & Hyönä, 2005; Poynor & Morris, 2003; Siyanova-Chanturia, Conklin, & Schmitt, 2011). Eye-tracking is useful in reading research for the following reasons. First, it allows participants to read a passage in a more natural situation by presenting an entire passage and not giving participants extra requirements (e.g., frequently pressing a button). Second, it discriminates between effects on early and late comprehension processes during reading. Third, it captures participants' regressions or look backs to the previously processed context. In sum, eye-tracking measures allow for real-time assessment of text processing under more natural conditions and during uninterrupted comprehension.

Fixation times and counts within the region of interest (e.g., word, phrase, or sentence) are often reported in eye-tracking studies. Furthermore, measures are commonly classified as early processing measures (e.g., first fixation duration and first-pass reading time) or late processing measures (e.g., total reading time and second-pass reading time). The former measures are sensitive to early processes during text comprehension, such as lexical access and early integration of text information, whereas the latter measures are sensitive to later processes during text comprehension, such as text reanalysis, discourse integration, and recovery from processing difficulties.

Eye tracking has been successfully adopted in studies of lexical processing and syntactic parsing. Nevertheless, Hyönä, Lorch, and Rinck (2003) suggested that eye tracking should be applied to the study of discourse processing. Similarly, Rayner, Chace, Slattery, and Ashby (2006) noted “the time is ripe for more comprehension studies to use eye movement data to understand discourse processing” (p. 252). Indeed, some studies have investigated eye movements during discourse processing. Most of these studies defined the clause or sentence, rather than the single word or phrase, as the unit of analysis (e.g., Kaakinen, Hyönä, & Keenan, 2002; Rinck et al., 2003; van der Schoot et al., 2012).

Calvo et al. (2001) used eye tracking to validate and extend prior findings on predictive inference generation that were obtained with other measures, such as naming or reading times. Using similar materials as Calvo and Castillo (1996, 1998), Calvo et al. found that second-pass reading time was facilitated and there was a reduction in regressions to target words representing a predictable event after reading the predictive context. Furthermore, there was an increase in regressions for words representing a non-predictable event. These results suggest that predictive inferences are generated on-line, but are active at late comprehension stages, rather than during early lexical access processes.

Over the last few years, eye tracking has also been used to investigate L2 acquisition and processing (see Dussias, 2010; Roberts & Siyanova-Chanturia, 2013, for a review). For example, Kadota and Kuramoto (2004) found that the number of fixations and regressions provided a sensitive measure for differentiating processing of garden-path (GP) and the non-GP sentences among Japanese EFL readers. In addition, it is generally assumed that L2 readers have longer fixation times and more fixations than L1 readers (mean fixation times on individual words are about 200–250 ms for skilled L1 readers of alphabetic languages; Rayner, 1998, 2009). Indeed, Nishiyama and Kadota (2004) reported that mean fixation times per word were around 330 ms in Japanese EFL readers. However, to my knowledge, eye tracking has not been used to investigate inferential processing in L2 reading.

2.3.3 Revising Predictive Inferences

As reviewed above, previous studies have enhanced our understanding of predictive inference generation during reading. However, readers do not always make correct predictive inferences. Consider that the aforementioned example text of throwing a vase (see Table 2.3) 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 et al., 1988). In this case, the readers' initial prediction of the vase breaking is not supported; therefore, it is necessary to revise the prediction to achieve appropriate comprehension of the text.

According to Rapp and Kendeou (2007), the term *revision* refers to a specific form of text comprehension updating that requires readers to replace incorrect or old information with the correct or new information. Incorrect inferences that are not successfully revised by readers lead to misunderstanding of the text or comprehension failures. Thus, it is important to investigate whether and how initially drawn inferences are revised when they are disconfirmed by the subsequent context. Despite its importance, however, in previous studies much less attention has been paid to the revision of predictive inferences. The following sections describe two cognitive mechanisms that are assumed to be necessary for revising inferences.

Suppression of inferences

To achieve successful inference revision, it is first necessary to suppress inference activation. As described in 2.1.1, suppression is defined as decreased activation of text information that is inappropriate or less relevant to current comprehension (Gernsbacher, 1990, 1997). This suppression mechanism allows readers to control attention resources effectively and flexibly during text processing. The suppression of explicitly stated text information, such as information about narrative characters and their goals, has been examined in some studies (e.g., Gernsbacher et al., 2004; Linderholm et al., 2004; see 2.1.1). However, the suppression of inferential information has been investigated in only a limited amount of research.

Potts et al. (1988), for example, examined the suppression of predictive inferences of L1 English readers. Participants read (a) predictive texts, in which target predictive inferences are suggested (e.g., *No longer able to control his anger, he threw a delicate porcelain vase against the wall*), (b) disconfirming texts, in which suggested inferences are disconfirmed by the subsequent context (e.g., *No longer…By sheer luck, the vase hit at an angle such that it was not damaged*), and (c) control texts, in which words that were semantically related to target inferences were included, but the inferences were not suggested (e.g., *In one final attempt to win the delicate porcelain vase, the angry husband threw the ball at the bowling pins that stood against the wall*). Lexical decision and naming times to target words (e.g., *break*) were collected and compared between the text conditions. Potts et al. supposed that if the disconfirming context serves to suppress the activation of initially drawn inferences, then reaction times in the disconfirming condition should not be facilitated due to lack of inference activation; reaction times should be longer than those in the predictive condition and the same as those in the control condition. If, on the other hand, the activation remains even after the inference is disconfirmed, then reaction times in the disconfirming condition should be facilitated by the activation; reaction times should be the same as those in the predictive condition and faster than those in the control condition. The results indicated no significant differences in reaction times between the predictive and disconfirming text conditions, suggesting that inference activation was not suppressed even after processing the context disconfirming the inferences.

Iseki (2006) conducted a similar experiment but yielded a result opposite to that of Potts et al. (1988); the activation of predictive inferences was suppressed during reading. In this study, L1 Japanese readers read predictive, disconfirming, and control texts and then performed the MJT. The analysis of judgment times indicated longer latencies for disconfirming texts than predictive texts, suggesting the suppression of predictive inferences after the disconfirming context. Thus, there is no converging evidence regarding the suppression of predictive inferences during reading.

Deletion of inferences

In addition to suppression, readers must delete incorrect inferences from long-term text memory to complete inference revision. Tapiero (2007) noted that the suppression and deletion should be considered as different processes: It is “worthwhile to assume that inhibition or suppression takes place when activated information is not necessary or relevant to the information currently being processed, and that a distinct deletion process takes place when information is eliminated from the episodic text representation” (p. 89).

Guéraud, Blanc, and Tapiero’s (2001) study (as cited in Tapiero, 2007) demonstrated that there are two mechanisms, suppression and deletion, involved in resolving narrative inconsistencies. In this study, participants performed a word-probe recognition task right after the display of the inconsistency, and answered an inference-making task after reading each passage. The results suggested that the contradictory information was suppressed and then reactivated when the passage provided an explanation to resolve the inconsistency. However, the contradictory information was deleted when the passage strengthened the inconsistency. Thus, the suppressed information may be reintegrated into or deleted from the readers' text representations depending on the subsequent text information (Tapiero, 2007). Based on these studies, in the current dissertation, suppression and deletion are assumed to be distinct cognitive processes in reading, and deletion of inferences is defined as elimination of disconfirmed inferential information from long-term text memory.

Some L1 studies examined the deletion of disconfirmed inferences in reading. For example, in Wilkes and Leatherbarrow (1988), participants read article reports in which the earlier-stated information about implicit causes of an event was disconfirmed and corrected by the subsequent information. The analysis of the written recall protocols and inferential questions found that incorrect inferences based on the early-stated information remained in participants' text memory, although they had clearly recalled that it had been disconfirmed and corrected. Similar results were obtained in Johnson and Seifert (1994). Thus, these studies indicated difficulties in editing text memory–failing to delete incorrect inferences in reading. However, both of these studies further revealed that deleting incorrect inferences could be successful if the earlier-stated information did not play an important role in text plots or if a plausible explanation for deleting was provided.

Although these past studies offer insights into deletion mechanism in reading, they dealt with inferences about implicit causes of narrative events during L1 reading. Few similar studies have been conducted to examine deletion of inferences about predictable narrative events in L2 reading.

Effects of disconfirmation of predictive inferences on explicit text comprehension

Although drawing predictive inferences during reading is beneficial for comprehension, Fincher-Kiefer (1993) noted that the risk of generating inferences that are inconsistent with the subsequent context is prohibitive because correcting text comprehension may be cognitively demanding. However, few empirical studies have examined the effects of disconfirmation of predictive inferences on comprehending explicit text information.

Nahatame (2010) is one of few studies that showed effects of disconfirmation of predictive inferences on reading comprehension. In this study, Japanese high school EFL students read short narrative passages and were asked to make predictions at a certain point during reading. Results of a written recall task after reading showed that recall performance was lower when readers made predictions that were inconsistent with the subsequent context and L2 proficiency was low. Similarly, Nahatame's (2011) study of Japanese university EFL students suggested that making predictions that are inconsistent with the following context increased the processing difficulty of the context. Thus, these results suggested negative effects of making incorrect predictive inferences on comprehension of explicit text information in L2 readers. However, few studies have addressed this issue. Therefore, more evidence is required to confirm and further investigate the effects of disconfirming predictive inferences on explicit comprehension in L2 readers.

2.4 Limitations of Past Studies and Connections to the Current Research

As reviewed so far, research and theoretical models of reading suggest that making inferences plays a significant role in reading comprehension. Therefore, a considerable number of studies have investigated inference generation, in particular predictive inferences, in L1 reading. In contrast, as Koda (2005) noted, only a limited number of studies have examined inference generation in L2 reading (e.g., Horiba, 1996; Muramoto, 2000; Yoshida, 2003). Furthermore, there are some limitations and unresolved issues in these studies.

First, as mentioned in 2.2.3, few L2 studies have focused on predictive inference generation in reading. Therefore, it is necessary to closely examine predictive inferences in L2 reading and identify the factors affecting predictive inference generation. Investigating predictive inferences in L2 reading is both theoretically and educationally important because predictive inferences have received considerable attention from reading researchers, and readers benefit from making these inferences (see 2.3.1).

Second, some previous studies tested inference generation in L2 reading with think-aloud tasks (Horiba, 1996, 2000; Yoshida, 2003), whereas most L1 studies assessed inference generation based on reaction times for target probe words (e.g., recognition times or lexical decision times). As described in 2.3.2.2, in think-aloud tasks participants are instructed to report whatever they are thinking about during reading. Therefore, although previous studies have reported some predictive inference generation during L2 reading, it is unknown whether these inferences are made in the absence of such instructions. Additionally, it is difficult to compare the results obtained in these L2 studies with the findings from the large body of L1 studies due to the different methodologies. Thus, it is important to examine predictive inference generation in L2 reading using methods other than the think-aloud task, such as a response time task.

Third, compared to predictive inference generation, only a few previous studies have investigated predictive inference revision in reading; unfortunately, the results are contradictory or unclear (e.g., Iseki, 2006; Potts et al., 1988). If incorrect predictive inferences are drawn, revising these inferences based on the subsequent context is necessary for successful comprehension. Therefore, this is an important issue that should be addressed.

Study 1 of this dissertation was conducted to address the first two issues mentioned above. This study investigated whether and how predictive inferences are made among Japanese EFL readers. Although previous studies suggest that several factors affect predictive inference generation during reading (see 2.3.2.1), it is difficult to control all of these factors in a few experiments. Consequently, this study examined the effects of specific text, reader, and task factors on predictive inference generation; specifically, contextual constraint and relatedness to narrative characters' motivation or goals (text factors), L2 reading proficiency and the amount of cognitive resources available during reading[2] (reader factors), and strategy and task instructions (task factors).

As described in 2.3.2.2, previous studies have used a variety of measures to assess predictive inference generation in reading. Although Study 1 employed several measures based on the specific purpose of each experiment, both on-line and off-line measures were used in all experiments. Horiba (2013) suggested that combining on-line and off-line measures provides a better picture of complex L2 text comprehension.

Study 2 in this dissertation was conducted to address the third issue mentioned above. The aim of this study was to examine whether and how incorrect predictive inferences are revised by Japanese EFL readers and how this influences reading comprehension, taking L2 reading proficiency into account. The experiments based on the results from Study 1. Each experiment in Study 2 also used both on-line and off-line measures.

[2] In Study 1, the amount of cognitive resources available during reading was manipulated within participants. Therefore, it is not an actual individual difference. However, it is likely that manipulating the amount of available cognitive resources simulates individual differences in readers' WM capacity.

Chapter 3

Study 1: Making Predictive Inferences in Japanese EFL Learners' Reading Comprehension

3.1 Experiment 1: Predictive Inference Generation and Text Characteristics in EFL Reading

3.1.1 Purpose and Research Questions

As reviewed in the previous chapter, several studies have suggested that text characteristics are a determinant factor affecting predictive inference generation during L1 reading. According to theories of inference generation (Graesser et al., 1994; McKoon & Ratcliff, 1992) and empirical studies (Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999), whether the inference relates to the narrative character's goal or motivation (i.e., whether inference is necessary for maintaining local coherence of the passage) is assumed to be one of the significant text factors for predictive inference generation. In addition, among the various text factors (see 2.3.2.1), the effects of contextual constraint on predictive inference generation have been extensively investigated in previous studies (e.g., Cook et al., 2001; Klin, Murray, et al., 1999; Linderholm, 2002; Virtue et al., 2006).

Although previous studies have suggested that L2 learners sometimes make predictive inferences during reading (Horiba, 1996; Yoshida, 2003), few studies have examined under what condition they are more likely to generate these inferences in terms of text characteristics. Therefore, Experiment 1 was conducted to examine predictive inference generation in Japanese EFL learners' reading comprehension, including the two aforementioned text factors: (a) whether the inference relates to the narrative character's goal or motivation (i.e., subtypes of predictive inferences: motivational or consequence inference) and (b) whether the context of the passage strongly constrains the possible inference (i.e., contextual constraint). The following two research questions (RQs) were addressed in this experiment:

RQ1-1: Do Japanese EFL readers make predictive inferences differently depending on subtypes of the inferences?

RQ1-2: Do Japanese EFL readers make predictive inferences differently depending on the contextual constraint?

This experiment used both on-line and off-line measures of predictive inference generation in reading. In this experiment, the word-probe recognition task was adopted as an on-line measure of inference generation, and recognition times to target probe words were analyzed.[3] As assumed in some previous studies (e.g., Fincher-Kiefer, 1995, 1996; McKoon & Ratcliff, 1986), a delay in recognizing an inferential target word is regarded as evidence of inference generation during reading (see 2.3.2.2 for a review).

Cued recall was used as an off-line measure of inferences in this experiment. In previous studies of predictive inferences, this has been one of the most frequently used tasks (e.g., Klin, Murray, et al., 1999; Murray & Burke, 2003; see also 2.3.2.2 for detailed explanation). In this task, the first sentence of each experimental passage was given to participants as a retrieval cue. If the predictive inference is drawn and becomes part of the long-term text memory, participants should falsely produce inferential information in the recall of subsequent text; on the other hand, if the inference is not drawn, there should be very few inference intrusions in recall (see also 2.3.2.2 for a review).

[3] Although previous studies have often used the lexical decision task to measure on-line predictive inference generation during L1 reading, obtaining reliable priming effects on L2 lexical decisions depends on participants' L2 proficiency (see 2.3.2.2, p. 48). Because it was unclear whether the task can be appropriately conducted with intermediate-level EFL learners (i.e., participants in this study), it seemed plausible to first adopt the word-probe recognition task and then use the lexical decision task to confirm the previous results. The word-probe recognition task simply requires participants to decide whether the target word appeared in the text, and therefore, it is assumed that even intermediate-level EFL learners can easily and appropriately perform the task.

3.1.2 Method

3.1.2.1 Participants

The participants were 30 Japanese undergraduate students (19 female and 11 male; aged 18–22 years, M = 19.70, SD = 0.99). They were in different years (first to fourth) and had different majors (e.g., pedagogy, physics, etc.). All participants had studied English as a foreign language for more than six years as part of formal Japanese education and they were assumed to have intermediate-level English proficiency.

Any data from one participant with more than 28% errors in the comprehension questions was excluded. Therefore, the analyses were based on data from 29 participants.

3.1.2.2 Materials

The materials consisted of (a) short narrative passages, (b) target words, and (c) comprehension questions. The passages and target words were based on those in Virtue et al. (2006) and Motyka Joss (2010).

Passages

Firstly, 96 short narrative passages written in English were adopted as candidate materials. These included 48 *inference* texts and 48 *neutral* texts. Each text consisted of four sentences. As these texts had been designed for native English-speaking readers, sentence structure was simplified and low-frequency words were replaced with high-frequency synonyms (e.g., *fleeting* → *running*, *stroll* → *walk*) so that EFL learners were able to understand text meaning. This process was conducted by referring to the *JACET List of 8000 Basic Words* (JACET Committee of Revising the JACET Basic Words, 2003). All the revised passages were checked and corrected by a native English speaker.

The inference texts were designed to elicit specific predictions for the outcomes of events described therein. Each inference text had two constraint versions, *strong constraint* and *weak constraint*. These two versions were introduced with the same three sentences, but were different in the fourth (last) sentence of the text, which either strongly or weakly constrained possible predictive inferences. The neutral texts were similar in length to inference texts. However, these texts described different topics from inference texts, and were unlikely to induce target predictions. These neutral texts were used to provide a baseline measure of inference activation.

Before the experiment, the inference texts were classified into two subcategories by three raters individually. The raters were graduate students majoring in applied linguistics, and the categories chosen depended on which type of predictive inference the passage induced—a motivational or consequence inference. Of 48 inference texts, 38 texts (19 motivational and 19 consequence inference passages) were chosen on which there was consensus among the three raters.

These inference texts were pilot tested to verify the probability of target predictive inferences based on the procedure of Campion and Rossi's (2001) preliminary study. All inference passages (strong constraint version) were presented to four university students majoring in applied linguistics. They read each text and then were asked to write an English verb representing a likely outcome of the event described in the text. Afterwards, the participants were presented with the actual target words for each inference passage (e.g., *steal* or *break*). Then they were required to evaluate the probability of that outcome represented by each target word on a 7-point scale ranging from 1 (*minimal probability*) to 7 (*maximum probability*). As a result of this pilot study, 32 inference passages (16 motivational and 16 consequence inference passages) were chosen as experimental inference passages due to their ease of guessing target words and high probability of suggested inferences.[4] Thirty-two

[4] The mean probability rate was not significantly different between selected motivational inference passages (M = 6.18, SD = 0.51) and consequence inference passages (M = 5.94, SD = 0.71), $t(30) = 1.12$, p

neutral passages corresponding to these inference passages were also selected. First, Table 3.1 shows examples of selected inference passages in each inference type and each constraint version. Second, Table 3.2 shows an example of neutral passages. Finally, Table 3.3 shows descriptive statistics for the experimental passages.

Table 3.1

Sample of the Inference Passages, Target Words, and Comprehension Questions in Experiment 1

Motivational inference text

Brad was looking for a present for his wife's birthday. He wanted to find something special for her, but he couldn't afford to buy anything nice. In the accessories department, he saw an expensive ruby ring sitting in a display. Seeing no salespeople or customers around, he quietly made his way to the display and opened his bag. (Strong constraint) / examined it more closely. (Weak constraint)

Target word: *steal*

Question: Was Brad looking for a popular book? (No)

Consequence inference text

Henry was a very careless man. He rarely watched where he was going. Today he was in a hurry to get home. As he was heading home, he stepped on some ice. (Strong constraint) / in some mud. (Weak constraint)

Target word: *slip*

Question: Was Henry very careless? (Yes)

= .273, *d* = 0.39.

Table 3.2

Sample of the Neutral Passages, Target Words, and Comprehension Questions in Experiment 1

Brad and Fred had been working in New York City for six months. Brad's parents were coming to stay with them this weekend. He was trying to think about what fun tourist spots they would visit. Brad and Fred decided to take them to the Museum of Art first.

Target word: *steal*

Question: *Were Brad's parents coming to Tokyo?* (No)

The woman entered the hall. Her reputation preceded her wherever she went. She was currently dating a handsome man half her age. Whispers followed her as she moved across the room.

Target word: *slip*

Question: *Was the woman dating a handsome man?* (Yes)

Table 3.3

Number of Words and the Readability of Experimental Passages in Experiment 1

		Number of words				Readability	
		M	*SD*	*Max*	*Min*	FKGL	FRE
Motivational	SC	49.00	7.54	61	36	4.98	80.81
	WC	48.19	7.52	59	34	5.11	79.58
Consequence	SC	50.00	9.16	67	32	5.36	79.34
	WC	49.44	9.14	62	32	5.19	79.75
Neutral		41.56	5.93	53	31	4.54	80.62

Note. SC = strong constraint; WC = weak constraint. FKGL = mean Flesch-Kincaid Grade Level; FRE = mean Flesch Reading Ease.

In addition to experimental passages, 16 filler texts were also used. The filler texts were similar in length to the experimental texts, but they did not elicit any specific inferences. All of the passages (both experimental and filler passages) as well as corresponding target words and comprehension questions used in this experiment are shown in Appendices 1 to 6.

Target words

Each experimental passage was paired with a corresponding target word for the recognition task. Target words were one- or two-syllable action verbs that represented the predictive inference concept suggested by the inference text (see Tables 3.1 and 3.2 for examples). These words were three to seven letters in length, and most of them appeared in the most frequent 3,000-word level (Levels 1 and 3) in the *JACET List of 8000 Basic Words*, and had higher familiarity ratings on a 7-point scale ranging from 1 (*least familiar*) to 7 (*most familiar*) in Yokokawa (2006). Table 3.4 shows descriptive statistics for the target words as a function of inference types. All the target words used in Experiment 1 are presented with their profiles (e.g., part of speech, frequency, and familiarity) in Appendix 8. There were no significant differences in the length, frequency, and familiarity of target words between motivational and consequence inference passages (all Fs < 1). In addition, there were no significant differences between four material sets (all Fs < 1.04).

To balance the responses in the recognition task, all target words paired with the filler texts were a simple word included in each filler passage. The target words for fillers were verbs, nouns, or adjectives (e.g., *lose*, *actor*, *shy*) in order to avoid attention being directed to only the verbs included in the passages.

Table 3.4

Length, Frequency, and Familiarity of Target Words in Experiment 1

	Motivational		Consequence	
	M	*SD*	*M*	*SD*
Length (number of letters)	4.56	1.03	4.38	0.89
Frequency (JACET 8000 Level)	1.75	1.81	1.81	1.72
Familiarity (on a 7-point scale from Yokokawa, 2006)	5.61	1.37	5.18	1.31

Note. There were two target words that were not included in either the *JACET List of the 8000 Basic Words* or Yokokawa's (2006) list of familiarity ratings (*skate* for a motivational inference text, and *sneeze* for a consequence inference text). These words were given minimum rates of frequency (eight) and familiarity (one).

Comprehension questions

A simple yes/no comprehension question was constructed for each passage to ensure that participants engaged in text reading (see Tables 3.1 and 3.2 for examples). These questions concerned an explicitly stated piece of information and were carefully written to be unrelated to the targeted inferences.

Four material sets were created resulting from the counterbalance of the experimental conditions using a Latin square. Each set included 24 of 32 experimental passages and all the filler passages (i.e., a total number of the passages was 40). This ensured that each participant read an equal number of passages in each of the six conditions (i.e., Motivational × Strong Constraint [MSC], Motivational × Weak Constraint [MWC], Motivational × Neutral [MN], Consequence × Strong Constraint [CSC], Consequence × Weak Constraint [CWC], and Consequence × Neutral [CN]), and each experimental passage was presented to one fourth of the participants in each condition.

3.1.2.3 Procedure

Participants were tested individually in a session that lasted approximately 70 minutes. The experiment had the following two phases: an on-line phase and an off-line phase. The procedures in these two phases are described in detail below.

On-line phase

The on-line phase employed the program SuperLab 4.0 (Cedrus, U.S.) on a computer. During this phase, participants were assigned to one of the four material sets. They read the instructions on the computer screen. The instructions informed them that their task involved reading a series of short passages, with each passage followed by a single-word recognition task, and then a yes/no comprehension question. Figure 3.1 represents the sequence of events during a trial in the on-line phase.

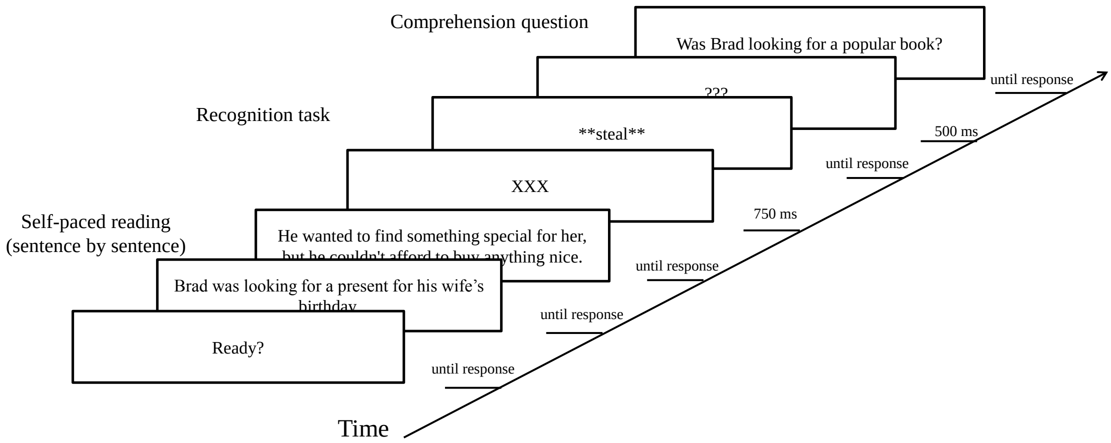

Figure 3.1. The sequence of events in each trial during the on-line phase of Experiment 1.

Each trial began with the word “Ready?” presented in the center of the computer screen. When ready to begin the trial, participants pushed the “yes” button on the Response Pad

RB-730 (Cedrus, U.S.) and the first sentence of the passage appeared. Participants read the text in a self-paced manner and controlled the pacing of presentation by pressing the "yes" button. Pushing the "yes" button removed the current sentence and replaced it with the next one. After reading the final sentence of the passage, pressing the "yes" button led to the appearance of a warning signal (XXX) for 750 ms.[5] The target word flanked by asterisks (e.g., **break**) followed the warning signal. Participants were required to determine, as quickly and accurately as possible, if the target word had occurred in the passage they had just read (i.e., a probe recognition task). They pressed the yes/no button to indicate their response. After responding, another signal (???) appeared for 500 ms and was replaced by a comprehension question. Participants responded again using a pair of yes/no buttons. After their responses, the question was erased from the screen and then immediate feedback concerning accuracy was automatically presented on the screen. This trial was repeated for each of the 40 passages. The passages were randomly presented to the participants. The participants were given five practice items before beginning the experiment, and followed the same procedure as in the experiment.

Off-line phase

After completing the on-line phase, each participant was given another booklet consisting of 16 inference passages, which they had not read in the first phase. Participants were required to read the passages within 15 minutes. Each participant read an equal number of passages in each constraint version and inference type. After reading the passages, participants received a worksheet that displayed the first sentences of each passage. Then, they were instructed to recall and write down as much about the remaining part of the

[5] This time parameter was relatively long compared to some previous L1 studies (cf. 250 ms in Cook et al., 2001; Linderholm, 2002; 500 ms in McKoon & Ratcliff, 1986; Murray & Burke, 2003). This aimed to provide the participants with sufficient time to generate an inference and decrease wrap-up effects on recognition latencies.

passages as possible in Japanese, using the first sentence as a retrieval cue (i.e., a cued recall task). Participants were given an unlimited amount of time to complete the recall task. Note that cued recall was not conducted for the passages read during the on-line phase because the presentation of target words during the recognition task was likely to affect later recall of inferential information.

3.1.2.4 Scoring and Analysis

Prior to the main analysis, recognition and comprehension question data were eliminated if participants mistakenly skipped a sentence in the passage, because accuracy and latencies might be affected if a certain sentence had been missed. In addition, recognition times falling more than 2.5 standard deviations above the mean for each participant were excluded, along with those less than 100 ms. This resulted in the exclusion of 4.74% of the recognition data.

Inference activation scores (Graesser et al., 2001; Virtue et al., 2006) were calculated for each inference condition. In this experiment, the scores were calculated based on recognition times, instead of lexical decision times or naming times. As mentioned earlier, if readers activated the inferential information represented by the target word during reading, then a correct response (i.e., “No”) would be slow. Thus, the activation scores were obtained by subtracting a participant’s mean correct response time for neutral texts (i.e., the MN and CN conditions) from the mean time for inference texts (i.e., the MSC, MWC, CSC, and CWC conditions). For example, if a participant’s mean response time was 1,000 ms in the MN condition and 1,100 ms in the MSC condition, the mean inference activation score for the participant in the MSC condition would be 100 ms. Overall, mean activation scores were then calculated for each inference condition.

Recall protocols were scored in terms of the inferential information produced therein.

This information was the inference concepts represented by the target words (e.g., *steal, slip*), which were not actually mentioned in the text. When the target word was falsely included in a participant's recall protocol, it was regarded as the production of inferential information and was scored with one point. Thirty percent of recall protocols were marked by the two raters separately, resulting in an agreement ratio of 98.56%. After resolving disagreements, the remaining data were scored by one rater. All the scores in each condition were transformed into a proportion, dividing the scores by the total number of passages in each condition.

3.1.3 Results

3.1.3.1 Comprehension Questions

The participants correctly answered, on average, 93.01% of the comprehension questions for the experimental passages. This result was similar to that of previous studies with L1 readers (e.g., Campion & Rossi, 2001; Iseki, 2006) and supported the claim that participants engaged in text reading during the on-line phase. Table 3.5 shows the mean correct answer rates (%) of comprehension questions as a function of inference type and text type.

Table 3.5

Mean Accuracy (%) on the Comprehension Questions for Experimental Passages in Experiment 1

	Strong Constraint		Weak Constraint		Neutral	
Inference Type	*M*	*SD*	*M*	*SD*	*M*	*SD*
Motivational	92.24	15.09	95.40	12.32	91.38	15.35
Consequence	93.10	13.19	93.97	12.77	93.10	11.37

A 2 (Inference Type: motivational, consequence) × 3 (Text Type: strong constraint, weak constraint, neutral) analysis of variance (ANOVA) was conducted on these correct answer rates, with both Inference Type and Text Type as within-participants variables. The results indicated neither a significant main effect of Inference Type nor of Text Type (see Table 3.6). The interaction between Inference Type and Text Type was also not significant.

Table 3.6

Summary Table for Two-Way ANOVA of the Effects of Inference and Text Types on Mean Comprehension Question Accuracy in Experiment 1

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Inference Type (I)	0.00	1	0.00	0.05	.825	.00
Error (I)	0.36	28	0.01			
Text Type (T)	0.02	2	0.01	0.47	.626	.02
Error (T)	1.17	56	0.02			
I × T	0.01	2	0.00	0.25	.782	.01
Error (I × T)	0.88	56	0.02			
Total	2.43	145				

3.1.3.2 Recognition Times and Accuracy

Table 3.7 shows mean correct response times for the recognition task in each condition. Figure 3.2 displays inference activation scores calculated from these correct response times. A 2 (Inference Type: motivational, consequence) × 2 (Contextual Constraint: strong, weak) ANOVA was conducted on activation scores, with Inference Type and Contextual Constraint as within-participants variables. The results indicated that a main effect of Inference Type approached significance, but neither a main effect of Contextual Constraint nor the interaction between Inference Type and Contextual Constraint was significant or marginally significant (see Table 3.8). As shown in Figure 3.2, the activation scores were higher for the motivational inference condition (M = 86.56, SD = 314.50) than for the consequence inference condition (M = −24.38, SD = 224.76).

Table 3.7

Mean Correct Recognition Times (ms) of Target Words in Experiment 1

	Strong constraint		Weak constraint		Neutral	
Inference Type	*M*	*SD*	*M*	*SD*	*M*	*SD*
Motivational	1,308	483	1,216	351	1,175	340
Consequence	1,176	286	1,147	254	1,186	292

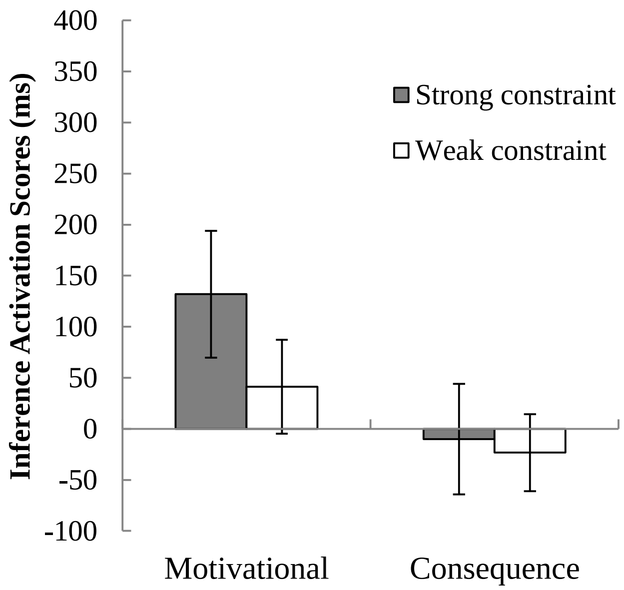

Figure 3.2. Mean inference activation scores (ms) as a function of inference type and contextual constraint. Error bars represent the standard errors of the means. MSC: M = 131.90, SD = 334.84; MWC: M = 41.21, SD = 291.50; CSC: M = –10.03, SD = 247.21; CWC: M = –23.38, SD = 221.32.

Table 3.8

Summary Table for Two-Way ANOVA of the Effects of Inference Type and Contextual Constraint on Inference Activation Scores in Experiment 1

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Inference Type (I)	356910.62	1	356910.62	4.04	.054	.13
Error (I)	2471597.53	28	88271.34			
Constraint (C)	103347.67	1	103347.67	1.54	.225	.05
Error (C)	1879353.01	28	67119.75			
I × C	27845.96	1	27845.96	0.38	.545	.01
Error (I × C)	2077430.96	28	74193.96			
Total	6916485.75	87				

To further explore the effects of Inference Type and Contextual Constraint on predictive inference generation, one-sample *t* tests were conducted on the activation scores in each of the four conditions (i.e., MSC, MWC, CSC, and MWC conditions). These tests examined whether the activation scores in each condition were significantly above zero (Virtue et al., 2006). The results showed that mean inference activation scores in the MSC condition were significantly greater than zero and the size of the effect was small to medium,[6] $t(28) = 2.12$, $p = .043$, $d = 0.39$. However, the scores were not significantly greater than zero in the MWC condition, $t(28) = 0.76$, $p = .453$, $d = 0.14$. The scores were also not significantly greater than zero in the CSC and CWC conditions, $t(28) = -0.22$, $p = .829$, $d = 0.04$; $t(29) = -1.03$, $p = .574$, $d = 0.19$.

Some previous studies (e.g., McKoon & Ratcliff, 1986; Shears et al., 2007) would have also regarded the false recognition or error response rates (i.e., "yes" response rates for target

[6] In this dissertation, the effect size *d* was calculated based on Cohen (1988). Cohen labeled an effect size small if $d = .20$, middle if $d = .50$, and large if $d = .80$.

words) as evidence of inference generation. Table 3.9 presents the mean recognition accuracy (%) of target words in this experiment.

Table 3.9

Mean Recognition Accuracy (%) of Target Words in Experiment 1

	Strong constraint		Weak constraint		Neutral	
Inference Type	*M*	*SD*	*M*	*SD*	*M*	*SD*
Motivational	91.38	15.08	92.24	12.75	100.00	0.00
Consequence	96.55	8.62	91.95	13.75	99.14	4.56

A two–way ANOVA was conducted on accuracy inhibition effects (i.e., differences of response accuracy between inference and neutral texts; MSC: M = 8.62, SD = 15.35; MWC: M = 7.76, SD = 12.97; CSC: M = 2.59, SD = 10.23; CWC: M = 7.18, SD = 13.12) with Inference Type (motivational, consequence) and Contextual Constraint (strong, weak) as within-participants variables. The results showed that neither a main effect of Inference Type nor Contextual Constraint was significant. There was also no significant interaction effect between Inference Type and Contextual Constraint (see Table 3.10). Although the descriptive statistics suggest some inhibition effects on recognition accuracy in each condition, it should be noted that the overall mean recognition accuracy was quite high and close to the ceiling (M = 95.30%, SD = 5.19). Therefore, the present experiment did not regard the response accuracy data as the test of inference. Instead, high success rates on the recognition task suggest that correct recognition of target words was not demanding for the participants in this experiment, who have intermediate-level English proficiency.

Table 3.10

Summary Table for Two-Way ANOVA of the Effects of Inference Type and Contextual Constraint on Recognition Accuracy Inhibition in Experiment 1

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Inference Type (I)	0.03	1	0.03	1.81	.190	.06
Error (I)	0.49	28	0.02			
Constraint (C)	0.01	1	0.01	0.77	.387	.03
Error (C)	0.37	28	0.01			
I × C	0.02	1	0.02	1.61	.215	.05
Error (I × C)	0.38	28	0.01			
Total	1.30	87				

3.1.3.3 Cued Recall

Table 3.11 shows the mean proportion of target inferences produced in recall protocols. A 2 (Inference Type: motivational, consequence) × 2 (Contextual Constraint: strong, weak) ANOVA on the mean proportion of the inference productions indicated a significant main effect of Inference Type, and a significant main effect of Contextual Constraint (see Table 3.12). Target inferences were recalled when the inferences were motivational more often than when they were consequence. Furthermore, participants recalled more target inferences when the contextual constraint was strong than when it was weak.

More importantly, there was a significant interaction between Inference Type and Contextual Constraint. To examine the simple main effect of Inference Type, the means for motivational and consequence inference conditions were compared for two constraint conditions using the Bonferroni correction (i.e., adjusted *p* value < .025).[7] Participants

[7] After applying the Bonferroni correction for the number of comparisons, *p* value is $< \alpha / 2 = .05 / 2 = .025$.

Table 3.11

Mean Recall Rates (%) of Target Inferences in Experiment 1

	Strong constraint		Weak constraint	
	M	*SD*	*M*	*SD*
Motivational	22.41	21.49	3.45	8.77
Consequence	5.17	10.31	0.86	4.64

Table 3.12

Summary Table for Two-Way ANOVA of the Effects of Inference Type and Contextual Constraint on Recall Rates of Target Inferences in Experiment 1

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Inference Type (I)	0.29	1	0.29	27.23	.000	.49
Error (I)	0.29	28	0.01			
Constraint (C)	0.39	1	0.39	19.63	.000	.41
Error (C)	0.56	28	0.02			
I × C	0.16	1	0.16	10.32	.003	.27
Error (I × C)	0.42	28	0.02			
Total	2.11	87				

produced significantly more motivational inferences than consequence inferences in the strong constraint condition, $t(28) = 4.61$, $p < .001$, $d = 1.02$, but the difference was neither significant nor marginally significant in the weak constraint condition, $t(28) = 1.36$, $p = .184$, $d = 0.37$. Similarly, the means for strong constraint and weak constraint conditions were compared for two inference type conditions using the Bonferroni correction. Participants produced significantly more motivational inferences in the strong constraint condition than in the weak constraint condition, $t(28) = 4.30$, $p < .001$, $d = 1.16$, whereas the difference was neither significant nor marginally significant for consequence inferences, $t(28) = 1.98$, $p = .057$, $d = 0.54$. These results demonstrated that target inferences were recalled most frequently in the MSC condition.

3.1.4 Discussion

RQ1-1 Do Japanese EFL readers make predictive inferences differently depending on subtypes of the inferences?

The analyses of recognition times indicated a marginally significant effect of inference subtypes on activation scores, such that the scores were higher for motivational inferences than consequence inferences, regardless of the contextual constraint. This result suggests the possibility that predictive inferences were more likely to be made when the inferences were motivational than consequence. This possibility was confirmed by the significant main effect of inference subtypes on recall performance: Inference productions in recall protocols were significantly higher for motivational inferences than consequence inferences. Taken together, the results indicate that predictive inferences were more likely to be made during reading and encoded into learners' long-term text memory when the inferences related to characters' goals or motivation than when they did not. The effects of the subtypes of predictive inferences are further discussed in response to RQ1-2, including consideration of the effects of contextual constraint.

Mean answer rates for the comprehension questions in the non-orienting condition were above 90% in each inference type condition. In addition, there was no significant difference in the answer rates between inference type conditions. These results suggest that readers were able to comprehend the experimental passages, regardless of inference type. Therefore, the possibility that readers did not make consequence inferences due to poor comprehension of the text’s meaning was ruled out.

The finding that motivational inferences are more likely to be generated than consequence ones is consistent with the results of previous L1 studies (Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999; Murray et al., 1993). According to both the minimalist hypothesis (McKoon & Ratcliff, 1992) and the constructionist theory (Graesser et al., 1994), inferences necessary for maintaining the local coherence of the passage are easily and automatically generated during reading. Because motivational inferences are required for preserving the local coherence of the passage, they are more likely to be made than consequence inferences, which are not required to maintain the local coherence.

In addition, of the possible inferences produced during narrative reading, Graesser et al. (1994) attributed central importance to inferences about the goals of the narrative character. These inferences explain why the character took the actions based on his or her implicit goals described previously. Motivational inferences represent a character’s goal (e.g., *steal*) as a predictable consequence of text events, and connect it to his or her action described in the sentence being processed to explain the reason for the action. This function of motivational inferences is similar to that of goal inferences, though the direction in which motivational inferences are made is opposite to goal inferences: Motivational inferences are made forward, whereas typical goal inferences are made backward (see Figure 3.3). Thus, motivational inferences can be interpreted as forward goal inferences. Because the causal links between the character’s goals and actions play an important role in narrative comprehension (Iseki & Kawasaki, 2006; Trabasso & Willey, 2005), motivational inferences are necessarily produced during reading.

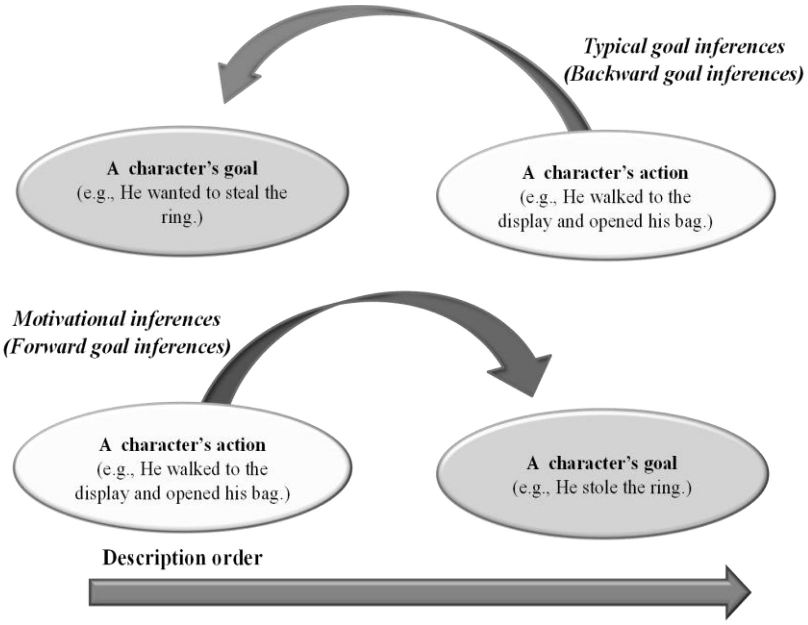

Figure 3.3. Backward and forward goal inferences.

As described above, theories in L1 reading research expect that motivational inferences are made more easily and necessarily during reading than consequence inferences. The present results suggest that this expectation can be applied to Japanese EFL learners' reading comprehension: Japanese EFL learners are also more likely to make motivational inferences than consequence inferences during reading.

RQ1-2 Do Japanese EFL readers make predictive inferences differently depending on the contextual constraint?

The results of the present experiment also suggest that motivational inferences were more likely to be generated when the contextual constraint was strong than when it was weak. The activation scores based on recognition times were significantly above zero in the MSC condition, but the effects were not significant in the other conditions (the MWC, CSC, and CWC conditions). Similarly, the results of cued recall indicated that there was a significant interaction effect between inference type and contextual constraint: Readers recalled target inferences to the greatest extent when motivational inferences were strongly constrained by the context, suggesting that target inferences were generated during reading and maintained in long-term text memory. Thus, the effects of contextual constraint suggested by L1 studies (Cook et al., 2001; Klin, Guzmán, et al., 1999; Linderholm, 2002) were also found in the EFL context, but only for motivational inferences. Similar to the discussion of RQ1-1, the high answer rates of comprehension questions ruled out the possibility that readers did not make inferences in the weak-constraint condition due to poor text comprehension.

As explained in previous studies, activation from a predictive sentence (i.e., the last sentence of the inference passage) might have been applied intensively to one possible and specific consequence when the contextual constraint was strong. In other words, when contextual constraint was strengthened, the information described in the sentence was sufficient to activate the inference concept represented by a target word. In contrast, when the contextual constraint was weak, the outcome of narrative events was less clear, which would result in no or insufficient activation of inference concepts. However, the effects of contextual constraint might be secondary to effects of subtypes of inferences because the results of the ANOVA on inference activation scores showed no significant or marginally significant main effect of contextual constraint, while the main effect of inference type approached significance.

In sum, the combination of the recognition time and cued recall results indicated the different likelihood of predictive inference generation among conditions: Predictive inferences were most likely to be made when the inferences were related to the narrative characters’ motivation and strongly constrained by context. Previous studies have demonstrated that L1 readers make predictive inferences during reading regardless of whether they are motivational or consequence inferences (Murray et al., 1993). However, consequence inferences are made only when the contextual constraint is strengthened (Klin, Guzmán, et al., 1999), while motivational inferences are made even when it is weakened (Klin, Murray, et al., 1999). Given these results, the results of this experiment suggest that predictive inference generation during EFL reading is more limited than during L1 reading. For example, this experiment found no evidence of predictive inference generation with consequence inferences, regardless of the strength of contextual constraint.

One possible explanation for the limited generation of predictive inferences in EFL reading is that EFL readers need to use a large proportion of their cognitive resources for lower level reading processing (e.g., word recognition and syntactic parsing), and therefore their higher level processing, which includes making inferences, is limited (Horiba, 1996; Yoshida, 2003). However, the experimental passages used in this experiment consisted of simple words and only four sentences, and the participants correctly answered most of the comprehension questions. This fact makes it unlikely that the participants in this experiment required considerable effort to comprehend the passages.

Another possible explanation is that EFL readers might engage in more careful reading than L1 readers. In other words, it is likely that EFL readers do not predict what will occur next during reading unless the prediction is necessary to achieve coherent comprehension and is strongly induced by a preceding context. This may be because readers try to avoid the risk of making incorrect predictions (Fincher-Kiefer, 1993), and adopt a wait-and-see strategy in which they wait until they have enough information to make highly probable predictions (Lassonde & O'Brien, 2009) for accurate comprehension (i.e., for correctly answering comprehension questions).

Further discussion of results obtained in Experiment 1

Experiment 1 regarded the activation scores as part of evidence of inference generation. Given that these scores represented the substantial difference in recognition times between inference and neutral texts, it could be argued that longer recognition times for motivational inference passages or shorter recognition times for neutral passages were caused by differences between these passages, rather than activating predictive inferences, as described below.

First, as shown in Table 3.3, inference passages were relatively long compared to neutral passages. This could increase memory load for an inference passage, resulting in a longer time to recognize the target words after reading the inference passage because recognition was performed based on readers' text memory. Second, inference passages included more words that were semantically related to the target words than the neutral texts (e.g., *afford, buy, expensive* vs. *steal*; *careless, hurry, step* vs. *slip*). This might make it difficult for readers to reject the target word due to activation of inter-lexical associations as a result of reading these inference-related words. On the other hand, because neutral passages described contents unrelated to the target words, it seems unlikely for participants to think about the target word while reading the passages.

However, these arguments cannot fully account for the obtained results. Again, it should be noted that activation scores were significantly greater than zero only for the MSC condition, but not for the other conditions (the MWC, CSC, CWC conditions). This occurred despite the fact that text length and target word relatedness to the text content did not differ between these conditions. Therefore, it is more probable that the results of the activation scores precisely reflect the strength of inference activation during reading, rather than mere differences in passage length and target word relatedness to text content.

In addition, regarding the results of cued recall, one may argue that mean recall rates of inferences were low even in the MSC condition (22.41%), and is therefore unlikely to be regarded as evidence for inference generation. However, it is important to note that recall of target inferences actually reflects an intrusion error: that is, recall of information that was not presented in the text. Thus, the production rates of inferences cannot be high. Indeed, the production rates in the MSC condition were close to those of previous L1 studies (approximately 23% in Klin, Murray, et al., 1999; approximately 30% in Murray & Burke, 2003).

3.1.5 Conclusion of Experiment 1

Experiment 1 investigated whether Japanese EFL learners make predictive inferences during reading, focusing on two text characteristics: subtypes of inferences (i.e., whether the inference relates to narrative characters' motivation) and contextual constraint. As described above, the combination of the recognition time and cued recall results demonstrated that predictive inferences were more likely to be made when the inferences related to narrative characters' motivation (i.e., motivational) than when they did not relate (i.e., consequence) (the answer to RQ1-1). In addition, the effects of contextual constraint were found only for motivational inferences: Motivational inferences were more likely to be made when then contextual constraint was strong than it was weak (the answer to RQ1-2). In sum, the present results suggest that Japanese EFL learners are most likely to generate these inferences while reading when the inferences are related to narrative characters' motivation and are strongly constrained by the context.

However, the on-line measure of inferences in this experiment (i.e., recognition times) showed minimal evidence of the effects of the aforementioned text factors on predictive inference generation. Jiang (2012) emphasized the importance of approaching a research topic or issue with different methods to refine the validity of the findings. In this way, Potts et al. (1988) conducted a series of experiments using lexical decision and naming tasks to examine predictive inference generation in L1 reading. Therefore, it is valuable to reexamine the findings of Experiment 1 (e.g., the high likelihood of inference generation under the MSC condition and the low likelihood of inference generation under the CSC condition) by using a different on-line measure than that used in Experiment 1.

One major problem with the word-probe recognition task is that it requires participants to access their text memory. Consequently, it is difficult to measure participants' inference generation implicitly. In contrast, the lexical decision task measures inference generation more implicitly than the word-probe recognition task because the task provides participants no clear reasons to access their text memory. Thus, the lexical decision task was used in the subsequent experiments, Experiments 2 and 3, to confirm the effects of text factors on predictive inference generation.

In addition, the aforementioned text factors may not always affect predictive inference generation in reading as expected under the influence of reader or task factors. As mentioned earlier, it is assumed that participants in the present experiment did not require much cognitive effort to comprehend the passages. However, both L1 and L2 studies suggest that inference generation during reading is limited when readers have fewer cognitive resources available for making inferences (Horiba, 1996; Linderholm, 2002; Yoshida, 2003). Thus, it is likely that EFL readers do not make predictive inferences under such a resource-limited condition, even though the text strongly induces the inference generation. The following experiment, Experiment 2, examined this possibility by manipulating the amount of participants' cognitive resources available during reading.

Additionally, Experiment 1 provided no evidence of the generation of consequence inferences. However, given the possibility that participants in Experiment 1 processed the passages carefully for accurate comprehension, consequence inferences might be generated during reading if the pre-reading instructions or post-reading tasks focused participants' attention on predictive inference generation. This possibility was investigated in Experiment 3.

3.2 Experiment 2: Predictive Inference Generation and Cognitive Demands in EFL Reading

3.2.1 Purpose, Hypothesis, and Research Question

The results of Experiment 1 suggested that Japanese EFL readers are more likely to make predictive inferences during reading when the passage strongly induces these inferences. However, as noted in 3.1.5, participants in Experiment 1 were unlikely to require much cognitive effort to comprehend the passages. Importantly, some L2 studies suggested that the amount of cognitive resources available during reading is a critical factor in determining inference generation (Horiba, 1996; Yoshida, 2003). However, these previous studies did not closely investigate the generation of predictive inferences. Therefore, it is necessary to directly investigate how predictive inference generation among L2 readers is affected by the amount of cognitive resources available during reading. The present experiment, Experiment 2, was conducted to address this issue.

The main variable in Experiment 2 was the degree of cognitive load given to participants. Participants were assigned to either *zero-load*, *low-load*, or *high-load* conditions while reading passages. To manipulate the degree of given cognitive load while reading, the present experiment employed Shears et al.'s (2007) dual-task methodology. This methodology requires participants to hold a list of words in memory while completing a reading task. Participants are then asked to recall the words after reading a number of passages (i.e., word-recall task). Such dual-task procedures divide participants' cognitive resources between memorizing words and reading passages. Therefore, the amount of cognitive resources available during reading decreases under dual-task conditions (i.e., low-load and high-load conditions) compared to a single-task condition (i.e., zero-load condition). The degree of cognitive load was also manipulated by varying the number of words to be memorized.

Given the findings of Experiment 1, it was hypothesized that when the passages strongly induce the inferences (i.e., when the inferences are related to narrative characters' motivation and are strongly constrained by the context), evidence of predictive inference generation would be obtained in the zero-load condition. Note that the zero-load condition is a replication of the MSC condition in Experiment 1. The main focus of this study was whether predictive inferences would be generated under two dual-task conditions (low- and high-load conditions). Thus, a hypothesis (H) and an RQ were addressed in Experiment 2 as follows:

H1: Japanese EFL learners make predictive inferences during reading when passages strongly induce the inferences and sufficient cognitive resources are available for making inferences.

RQ2: Do Japanese EFL readers make predictive inferences differently depending on the amount of cognitive resources available during reading?

Although Fincher-Kiefer and D'Agostino (2004) suggested that verbal working memory is not required to support predictive inference generation, Experiment 2 adopted a word-recall task as a means to increase cognitive demands on participants for the following reasons. First, Shears et al. (2007) contended that verbal working memory is required for all reading comprehension processes. Second, the word-recall task allows for easy variation of cognitive load given to participants by manipulating the number of words to be held in memory. This manipulation will provide a better understanding of the relationship between the amount of available cognitive resources and predictive inference generation.

Experiment 2 adopted a lexical decision task, which is one of the most frequently used tools to measure predictive inference generation in reading (e.g., Allbritton, 2004; Campion & Rossi, 2001; Virtue et al., 2006), rather than a word recognition task. The rationale for using the lexical decision task is that it provides participants no clear reason to compare a target word to their text memory. Consequently, the task measures inference generation more implicitly than the word-probe recognition task. Moreover, given that dual-task conditions increase participants' memory load and the recognition task requires participants to check their text memory, recognition times for the target words are assumed to be affected more strongly by memory load than inference generation. This might make it difficult to precisely assess inference generation during reading. Therefore, a task that does not require participants' judgments based on text memory seemed more suitable and appropriate in this experiment. If the results of the lexical decision task support H1, it strengthens the validity of the results obtained from the recognition task in Experiment 1.

Additionally, Experiment 2 employed probability judgment rather than cued recall as an off-line measure of inferences. When using a cued recall task, it is recommendable to have participants read new passages that differ from those used during the on-line task. As noted in Experiment 1, this is because the presentation of target words during the on-line task is likely to facilitate production of target words in later recall. However, it was difficult to prepare new passages that differed from those for the on-line task in this experiment because of the restriction of the experimental design and the number of experimental passages. In the probability judgment task, presenting target words during the on-line phase is less likely to affect task performance because the task does not assess recall of inferences: Participants are presented with target sentences describing possible future events and required to simply judge the probabilities of the events. Therefore, participants do not need to read new passages for making probability judgments.

3.2.2. Method

3.2.2.1 Participants

The participants were 36 Japanese undergraduate and graduate students (23 male and 13 female; aged 18–21 years, *M* = 18.83, *SD* = 0.94) from the same university. They were in different years (first to third) and had different majors (e.g., practical engineering, international studies etc.). All participants had studied English as a foreign language for more than six years as part of formal Japanese education and they were assumed to have intermediate-level English proficiency.

All data from four participants were excluded because two of them failed to follow the instructions and the others made errors on more than two thirds of the comprehension questions. Additionally, data from two participants were eliminated based on box plot extreme outlier detection in inference activation scores.[8] Therefore, data from 30 participants were included in the analyses.

3.2.2.2 Materials

Passages, target words, and comprehension questions

Twenty-eight short narrative passages were adapted from Virtue et al. (2006) and Motyka Joss (2010). The majority were the same experimental passages used in Experiment 1.[9] These included 14 *inference* passages and 14 *neutral* passages. The inference passages were the MSC (Motivational-Strong Constraint) version, which fulfilled the conditions for strongly inducing predictive inferences suggested by Experiment 1. The neutral passages were similar to those utilized in Experiment 1 (see Appendix 3). Table 3.13 shows descriptive statistics for the experimental passages used in Experiment 2. In addition to the experimental passages, 18 filler passages were used. The filler texts were similar in length to the experimental texts, but they did not elicit any specific inferences.

[8] Extreme outliers indicate values that more than three times the inter-quartile range of a box plot. This criterion was applied to all experiments in this dissertation.

[9] Because the lexical decision task was utilized in Experiment 2, it was necessary to control the number of letters, familiarity, and frequency of target words more strictly than in Experiment 1. Therefore, three inference passages in Experiment 1 whose target words had more letters or lower familiarity and frequency were excluded, and one new passage was added.

Table 3.13

Number of Words and the Readability of Experimental Passages in Experiment 2

	M	*SD*	*Max*	*Min*	FKGL	FRE
Inference	51.00	8.65	68	36	5.45	78.31
Neutral	41.00	7.58	62	30	3.79	85.34

Note. FKGL = mean Flesch-Kincaid Grade Level; FRE = mean Flesch Reading Ease.

As in Experiment 1, each passage had a corresponding target word that represented inferential information suggested by the inference text (e.g., *steal*). Because the lexical decision task was conducted in the present experiment, these target words were more carefully determined than in Experiment 1. As noted in 2.3.2.2, L2 lexical decisions are likely to be affected by participants' L2 proficiency and some word factors (e.g., frequency and familiarity). Thus, only highly frequent and familiar words were used as lexical decision probes for experimental passages so that intermediate level learners would have little difficulty accessing and processing them. All target words for experimental passages were three to five letters in length, appeared in the most frequent 3,000-word level (Levels 1 to 3) in the *JACET List of 8000 Basic Words* (JACET Committee of Basic Words Revision, 2003), and had familiarity ratings greater than four on Yokokawa's (2006) 7-point scale. Table 3.14 shows descriptive statistics for the target words used in Experiment 2 (see Appendix 8 for all the target words with their profiles). These factors (i.e., word length, frequency and familiarity) were counterbalanced across six material sets, and there were no significant differences between sets (all *F*s < 1.15). All target words paired with the filler texts were pseudowords to balance the responses in the lexical decision task. Pseudowords were created by rearranging the letters of real words so that they remained pronounceable (e.g., *clak*, *dal*).

Table 3.14

Length, Frequency, and Familiarity of Target Words in Experiment 2

	M	*SD*	*Max*	*Min*
Length (number of letters)	4.36	0.84	5	3
Frequency (JACET 8000 Level)	1.21	0.58	3	1
Familiarity (on a 7-point scale from Yokokawa, 2006)	5.87	0.64	6.77	4.27

Finally, each text was paired with a yes/no comprehension question as in Experiment 1. Comprehension questions focused on details of the passages and were used to encourage careful reading. These questions were not directly relevant to the target inferences.

Six material sets that included all the experimental and filler passages were created. Specific passages were counterbalanced across conditions using a Latin square to ensure that each participant read a similar number of passages in each of the six conditions (i.e., Inference-High Load, Inference-Low Load, Inference-Zero Load, Neutral-High Load, Neutral-Low Load, and Neutral-Zero Load) and did not read the same passage twice.

Word-recall lists

For the word-recall task, 27 words were selected from the sub-list 1 (i.e., the most frequent word level) of the new Academic Word List (AWL; Coxhead, 2000)[10] such that no chosen words were included in any of the experimental passages or were used as target words in the lexical decision task. Although Shears et al. (2007) conducted the word-recall task with native speakers of English, the present experiment was conducted with Japanese learners of English. Thus, only high-frequency words were included in the lists to reduce the task

[10] To create word-recall lists, it was necessary to select highly frequent words that did not appear in experimental passages. The new AWL was used to find such words because it included highly frequent academic words that are less likely to appear in the passages describing everyday topics.

difficulty. The selected words included verbs, nouns, and adjectives (see Appendix 10 for all the selected words with their profiles). Table 3.15 shows descriptive statistics for the words included in the lists.

Table 3.15

Length, Frequency, and Familiarity of Words Used in the Word-Recall Task in Experiment 2

	M	*SD*	*Max*	*Min*
Length (number of letters)	6.59	1.42	9	4
Frequency (JACET 8000 Level)	1.63	0.49	2	1
Familiarity (on a 7-point scale from Yokokawa, 2006)	4.82	0.70	6.23	3.63

Six word lists were created from the selected words. Three of the word lists contained three words (e.g., *legal, factor, theory*), and the other three word lists contained six words (e.g., *data, assume, benefit, percent, approach, identify*), respectively. The number of words included in the lists was determined based on the results of a small pilot study to ensure that participants could successfully perform both the word-recall task and the reading task at the same time. Mean number of letters, word frequency, and familiarity did not differ significantly between these six word lists (all Fs < 1).

Probability judgment sentences

In the probability judgment task, one judgment sentence was created for each inference passage, for a total of 14 judgment sentences. These sentences expressed specific predictable events suggested by the inference passage, and included the main character and the target word (e.g., *He stole the ring*). Each judgment sentence was paired with a probability scale that ranged from 1 (*minimal probability*) to 7 (*maximum probability*).

3.2.2.3 Procedure

Participants were individually tested in a session that lasted approximately 60 minutes. As in Experiment 1, the procedures of the session were divided into on-line and off-line phases. First, instructions at the beginning of the on-line phase informed participants that they would be required to perform two tasks at the same time. The word-recall task was described first, and participants were instructed to actively hold the listed words in memory while completing the reading task. Participants were also informed that they would read short narrative passages in English and make a series of yes/no responses to target words and comprehension questions. Participants were told that no word lists would be given on some trials, and that they were only to engage in the reading task on those trials.

On each trial during the on-line phase, participants followed the procedure shown in Figure 3.4. In the low-load and high-load conditions, participants pushed the "yes" button on the Response Pad RB-730 (Cedrus, U.S.), and then the word list was presented, along with a message instructing participants to memorize the presented words. In the high-load condition, six words were presented for 40 seconds, whereas three words appeared for 20 seconds in the low-load condition.[11] Next, the screen cleared and a message appeared telling participants to get ready to read the passage. In the zero-load condition a message was presented informing participants that there were no words to learn for that trial.

After word learning, the first sentence of the inference, neutral, or filler passage appeared in the center of the screen. Then, participants followed the same procedures as in Experiment 1, except that they performed a lexical decision task instead of a recognition task immediately after reading each passage. In the lexical decision tests, participants were required to determine if the presented target word was a real word or non-word as quickly and accurately as possible. Their accuracy and response latencies were recorded. The reading task

[11] The presentation time of each word list was also based on the results of a pilot study to ensure that participants had enough time to memorize the words.

was repeated for four passages that were presented to participants in random order.

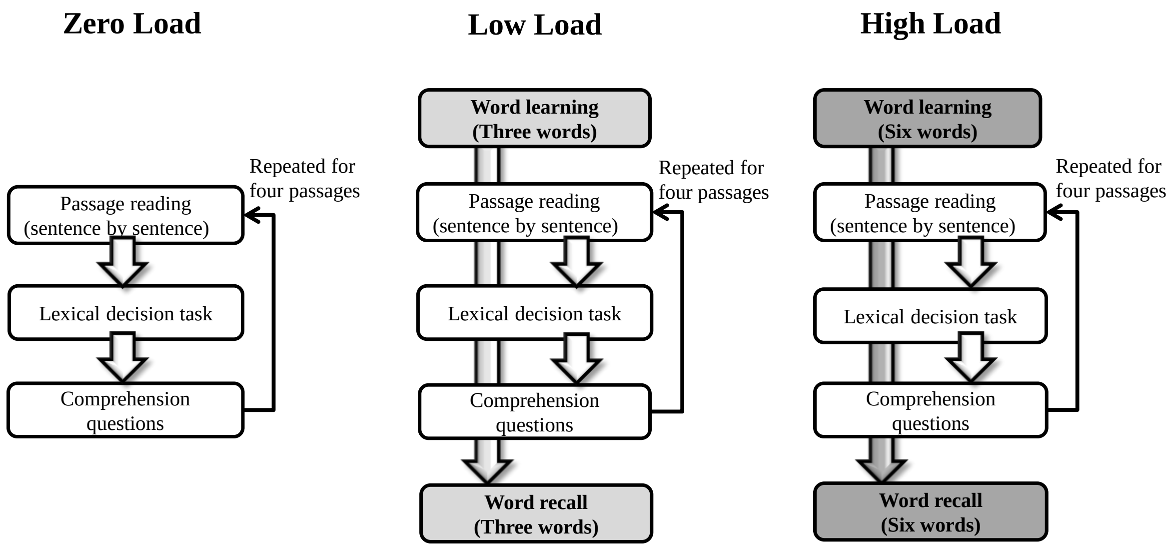

Figure 3.4. The sequence of events in each trial during the on-line phase of Experiment 2.

Following the reading task, a message appeared instructing participants to recall as many of the words they had studied during the word learning phase as possible. They wrote down the words on a sheet of paper. In the zero-load condition, a message was displayed telling participants to get ready for the next trial.

Each participant completed three trials for each load condition. The order of load conditions and text passages were counterbalanced across participants. To familiarize participants with the procedure, some practice trials were completed before beginning the experimental trials.

After completing all trials in the on-line phase, participants performed an unexpected probability judgment task in an off-line phase. Participants were presented with a booklet containing 14 judgment sentences with probability scales. Participants were instructed to decide if each sentence expressed an event that could happen next in the situation described by one of the passages they had read. If they answered “yes,” they had to evaluate the probability of the event by circling the appropriate number on a 7-point scale.

Although probability judgment was also conducted for neutral passages, data for these passages were not included in the analysis because they were unrelated to the target inferences expressed by the judgment sentences. The judgment sentences for the neutral passages were included simply to balance the number of yes/no responses in this task.

3.2.2.4 Scoring and Analysis

The word-recall task was scored using both the sensitive and strict scoring systems described by Webb (2008). In the sensitive scoring system, recalled words with spelling errors were judged as correct if the overall shape of the word is a close to the target word; in the strict scoring system, recalled words were only judged as correct if the target words were recalled with correct spelling. Recall scores for each trial were transformed into a proportion, dividing the number of correctly recalled words by the total number of words included in the list (i.e., three or six).

Participants’ lexical decision data as well as comprehension question data for a passage were eliminated if they accidentally skipped a sentence during reading. In addition, lexical decision times that were 2.5 standard deviations from the mean for a participant and those less than 100 ms were excluded. Together, this resulted in discarding 2.68% of the lexical decision data. As in Experiment 1, inference activation scores were calculated for each load condition based on mean correct lexical decision times. These scores were obtained by subtracting the mean response times in each inference condition (i.e., Inference-High, Inference-Low, Inference-Zero) from those in the corresponding neutral condition (i.e., Neutral-High, Neutral-Low, and Neutral-Zero). If the inference has just been activated, correct response times should be facilitated for inference contexts than for neutral contexts (see 2.3.2.2 for a review), resulting in activation scores greater than zero.

For the probability judgment task, mean proportions of positive responses (i.e., “yes” responses) were calculated for each inference condition, dividing the number of positive responses by the total number of judgment sentences in each condition. Additionally, mean probability ratings for judgment sentences were calculated. If the inference has been encoded in long-term text memory, participants should respond “yes” to judgment sentences and evaluate the sentences as highly probable.

3.2.3 Results

3.2.3.1 Word Recall

Table 3.16 shows mean word recall rates (%) in the high- and low-load conditions. In each condition, these rates were relatively high (approximately 65% or higher), indicating that participants allocated cognitive resources to holding the word lists in memory during reading.[12]

Table 3.16

Mean Word Recall Rates (%) in High- and Low-Load Conditions

	Strict scoring		Sensitive scoring	
	M	*SD*	*M*	*SD*
Low-load	83.70	12.96	87.78	12.50
High-load	65.37	15.89	68.52	16.46

Before addressing the question of inference generation, it is necessary to confirm the validity of the cognitive load manipulation. That is, word recall rates should be significantly

[12] Participants in Shears et al.’s (2007) study were able to recall approximately 95% of the words under low-load (i.e., three-word) condition and 80% of the words under high-load (i.e., six-word) condition.

higher in the low-load (i.e., three-word) condition than in the high-load (i.e., six-word) condition (Shears et al., 2007). Two-tailed paired *t* tests found that participants recalled words more accurately in the low-load condition than in the high-load condition according to both the strict and sensitive scoring systems, $t(29) = 5.25$, $p < .001$, $d = 1.27$; $t(29) = 5.60$, $p < .001$, $d = 1.32$.

3.2.3.2 Comprehension Questions

The participants correctly answered, on average, 92.33% of the comprehension questions for the experimental passages. Table 3.17 shows the mean correct response rates (%) of comprehension questions of experimental passages in each condition.

Table 3.17

Mean Accuracy (%) on the Comprehension Questions for Experimental Passages in Experiment 2

	Cognitive load					
	High		Low		Zero	
Text	*M*	*SD*	*M*	*SD*	*M*	*SD*
Inference	89.44	20.29	92.22	18.43	91.11	18.43
Neutral	95.56	13.79	88.89	19.25	93.89	16.07

A 2 (Text Type: inference, neutral) × 3 (Cognitive Load: high, low, zero) ANOVA was conducted on mean correct response rates with Text Type and Cognitive Load as within-participants variables. The results revealed no significant main effects of either Text Type or Cognitive Load. Nor was an interaction between Text Type and Cognitive Load significant (see Table 3.18).

Table 3.18

Summary Table for Two-Way ANOVA of the Effects of Text Type and Cognitive Load on Mean Comprehension Question Accuracy in Experiment 2

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Text Type (T)	0.02	1	0.02	0.60	.444	.02
Error (T)	0.74	29	0.03			
Cognitive Load (C)	0.02	2	0.01	0.19	.828	.01
Error (C)	2.31	58	0.04			
T × C	0.07	2	0.03	1.40	.254	.05
Error (T × C)	1.42	58	0.02			
Total	4.57	150				

3.2.3.3 Lexical Decision Times and Accuracy

Mean lexical decision times for correct responses are shown in Table 3.19. Overall mean response accuracy of lexical decision was 96.45%. This high success rate could be due to the use of highly frequent and familiar words as lexical decision probes, suggesting that these words were easily accessed and processed by the participants. Indeed, mean lexical decision times were close to those observed for native speakers of English in previous research using similar passages and target words (Virtue et al., 2006). Figure 3.5 displays inference activation scores calculated from correct response times as a function of cognitive load.

Table 3.19

Mean Correct Lexical Decision Times (ms) of Target Words in Experiment 2

	Cognitive load					
	High		Low		Zero	
Text	*M*	*SD*	*M*	*SD*	*M*	*SD*
Inference	793	279	734	195	745	184
Neutral	791	223	768	186	830	194

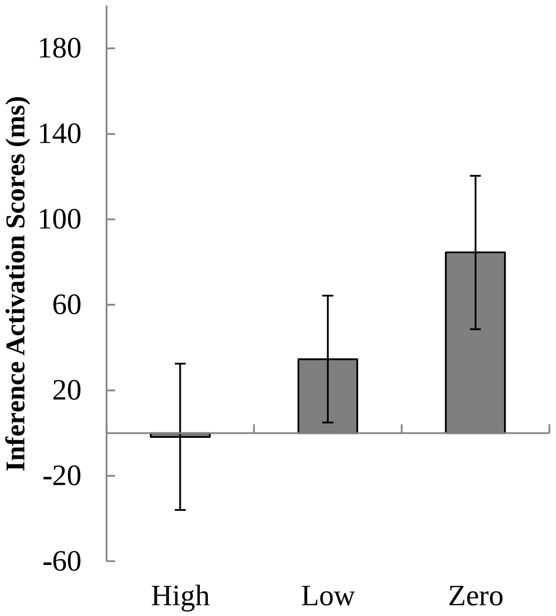

Figure 3.5. Mean inference activation scores (ms) as a function of cognitive load. Error bars represent the standard errors of the means. High: M = –1.76, SD = 187.62; Low: M = 34.62, SD = 162.35; Zero: M = 84.51, SD = 196.39.

A one-way ANOVA was conducted on the activation scores with Cognitive Load as a within-participants variable. Although the mean scores numerically increased from high load to zero load, the results did not indicate a significant effect of Cognitive Load (see Table 3.20).

That is, there was no significant difference between the three load conditions.

Table 3.20

Summary Table for One-Way ANOVA of the Effects of Cognitive Load on Inference Activation Scores in Experiment 2

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Cognitive Load	112555.45	2	56277.72	1.50	.231	.05
Error (Cognitive Load)	2170869.63	58	37428.79			
Total	2283425.07	60				

To further explore the effects of cognitive load on predictive inference activation, activation scores for each cognitive load condition were compared to baseline (i.e., zero). Two-tailed one sample *t* tests were conducted on the inference activation scores for each level of cognitive load to examine whether the activation scores in each load condition were significantly above zero. The activation scores was significantly greater than zero in the zero-load condition and the effect size was similar to that of Experiment 1, $t(29) = 2.36$, $p = .025$, $d = 0.42$. In contrast, these scores were not significantly greater than zero in the low-load condition and only a small effect size was observed, $t(29) = 1.17$, $p = .252$, $d = 0.21$. In the high-load condition, the activation scores were not greater than zero and the effect size was negligible, $t(29) = -0.51$, $p = .959$, $d = 0.01$.

Finally, facilitation of response accuracy on the lexical decision task was analyzed to test for a possible speed-accuracy trade-off (see Table 3.21 for mean response accuracy in each condition). A one-way ANOVA was conducted on facilitation effects (i.e., differences in response accuracy between inference and neutral texts) with Cognitive Load as a within-participant variable (High: $M = –2.22$, $SD = 16.80$; Low: $M = 2.22$, $SD = 22.20$; Zero:

M = 3.89, *SD* = 14.96). The result indicated that a main effect of Cognitive Load was not significant (see Table 3.22). As noted above, mean accuracy was very high in each condition, and therefore, there was no evidence of a trade-off between speed and accuracy in the lexical decision task.

Table 3.21

Mean Lexical Decision Accuracy (%) of Target Words in Experiment 2

	Cognitive load					
	High		Low		Zero	
Text	*M*	*SD*	*M*	*SD*	*M*	*SD*
Inference	95.56	13.79	96.67	12.69	100.00	0.00
Neutral	97.78	8.46	94.44	17.14	96.11	14.96

Table 3.22

Summary Table for One-Way ANOVA of the Effects of Cognitive Load on Facilitation of Lexical Decision Accuracy in Experiment 2

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Cognitive Load	0.06	2	0.03	0.92	.404	.03
Error (Cognitive Load)	1.88	58	0.03			
Total	1.94	60				

3.2.3.4 Probability Judgments

Table 3.23 shows the mean proportion of positive responses (%) and mean probability ratings in the judgment task. Separate one-way ANOVAs were conducted on mean positive response rates and mean probability ratings with Cognitive Load as a within-participants variable. A main effect of Cognitive Load was not significant for either the positive response rates or the probability ratings (see Tables 3.24 and 3.25). That is, probability judgments did not significantly differ between the three load conditions. As shown in Table 3.23, the mean percentage of positive responses and mean probability ratings were quite high in all conditions: The former was approximately 90% or more, and the latter was almost 6.00 or more.

Table 3.23

Mean Positive Response Rates (%) and Mean Probability Ratings in the Probability Judgment Task

	Cognitive load					
	High		Low		Zero	
	M	*SD*	*M*	*SD*	*M*	*SD*
Positive responses (%)	95.56	13.79	91.67	18.95	89.44	23.77
Probability ratings	5.99	1.00	6.39	0.61	6.31	0.88

Note. Probability was rated on a scale ranging from 1 (*minimal probability*) to 7 (*maximum probability*).

Table 3.24

Summary Table for One-Way ANOVA of the Effects of Cognitive Load on Mean Positive Response Rates in the Probability Judgment Task

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Cognitive Load	0.06	2.00	0.03	0.86	.430	.03
Error (Cognitive Load)	1.94	58.00	0.03			
Total	2.00	60.00				

Table 3.25

Summary Table for One-Way ANOVA of the Effects of Cognitive Load on Mean Probability Ratings in the Probability Judgment Task

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Cognitive Load	2.70	1.47	1.83	2.36	.120	.08
Error (Cognitive Load)	33.15	42.70	0.78			
Total	35.85	44.18				

Note. *df* were adjusted with Greenhouse-Geisser correction.

3.2.4 Discussion

Results of word recall and comprehension questions

Participants showed good performance both on the word recall task and comprehension questions. This suggests that participants divided their cognitive resources between holding a list of words in memory and reading passages under dual-task conditions. In addition, mean word recall rates were significantly lower in the high-load condition than the low-load condition. This confirmed the validity of the cognitive load manipulation in the experiment: More cognitive resources were required for high load than for low load.

The following sections discuss the effects of cognitive load manipulation on predictive inference generation during reading (RQ2) based on the results of the lexical decision task and the probability judgment task, respectively.

Lexical decision time results

First, *t* tests on inference activation scores, which were calculated from lexical decision times, demonstrated that these scores were significantly greater than zero in the zero-load condition with a small to medium effect size. This indicates that predictive inferences were activated during reading under the zero-load condition. This is consistent with the results of recognition times in Experiment 1, which found that activation scores were significantly above zero in the MSC condition and had a similar effect size. Thus, even when the more implicit measure of inferences (i.e., the lexical decision task) was used, the results indicated the effects of text characteristics on predictive inference generation during EFL reading: Japanese EFL readers are more likely to make predictive inferences when the inferences are related to narrative characters' motivation and are strongly constrained by the context. On the other hand, there was no evidence for making inferences under dual-task conditions. Specifically, the activation scores were not significantly greater than zero under these conditions. These results suggest that inference processing was impaired when the amount of cognitive resources available during reading were reduced. This is consistent with previous studies showing that cognitive demands during reading affect the inference generation among L2 readers (Horiba, 1996; Yoshida, 2003). The reason for impaired inference generation in this experiment can be explained in a similar manner as Horiba (1996) and Yoshida (2003): Holding a word list in WM reduced the cognitive resources available to support predictive inference generation. Furthermore, despite using passages that should strongly induce predictive inferences, the present results showed that predictive inference generation was less likely to be drawn not only under high load, but also under low load. This suggests that predictive inference generation was a process that requires more cognitive resources than were available in the low-load condition.

However, the results pertaining to lexical decision times should be interpreted with care because the ANOVA on inference activation scores indicated no significant differences between the three conditions. Taken together, the most plausible interpretation of the lexical decision time results is that increased cognitive demands were likely to impair predictive inference generation during reading, although the effect was relatively small (at least in this experiment). Such a small effect of increased cognitive demand may be explained in a number of ways. One possible reason could be that the demands elicited by the word-memory task were not higher than expected. Given that participants showed good performance on comprehension questions even under the high-load condition (the mean correct response rate was 92.56%), participants might not allocate many cognitive resources to holding a list of words in memory. Another possible reason is that, as pointed out by Fincher-Kiefer and D'Agostino (2004), demands on verbal working memory only had a small impact on the generation of predictive inferences compared to demands on visuospatial memory. Consequently, predictive inference generation might be more strongly affected by dual-task conditions if participants were given visuospatial memory load rather than verbal memory load.

Probability judgment results

In contrast to the results of lexical decision times, the analysis of the probability judgment task clearly indicated that target predictive inferences were evaluated as highly plausible in all load conditions. This suggests that readers generated these inferences and maintained them in long-term text memory regardless of the level of cognitive load. However, this interpretation is somewhat incompatible with those of the lexical decision times.

Given the difference in testing times between on-line and off-line tasks, the lexical decision and probability judgment tasks should be considered to measure different processes of inference generation in reading: The lexical decision task assessed inference activation during the course of reading, whereas the probability judgment task tested the maintenance of inferences in long-term memory after all the passages have been read. Horiba (2013) suggested that although the process of reading (assessed by the on-line task) often lead to the outcome of reading (assessed by the off-line task), there are some cases where the process is not reflected in the product. The dissociation between the results of the lexical decision and probability judgment tasks suggests the possibility that inferences were generated even under dual-task conditions, but they were drawn in different ways compared to the zero-load condition. Some possible mechanisms underlying inference generation are described below.

One possibility is that the instructions on the judgment task, which asked participants to determine and rate the probability of inferences, encouraged participants to draw inferences that were not activated during reading under dual-task conditions. In other words, participants might have been encouraged to generate inferences based on their text memory when they performed the judgment task (i.e., off-line generation of inferences), even though they did not automatically generate inferences when performing the lexical decision task (i.e., on-line generation of inferences).

Another possibility is that predictive inferences were drawn on-line even under dual-task conditions, but with some delay relative to the zero-load condition. More specifically, under dual-task conditions, readers may have required more than 750 ms (i.e., the time between the final passage sentence and the presentation of the target word) for inferences to reach a high enough activation level to be detected in the lexical decision task. This explanation assumes that inference activation was weak immediately after the lexical decision task, completed with some delay, and then maintained until the judgment task was performed. Linderholm (2002) suggested the possibility that low-WM capacity readers need more time to generate predictive inferences than high-WM capacity readers. Given this possibility, reducing cognitive resources in dual-task conditions might increase the time needed to complete inference generation, as with low-WM capacity readers. Consequently, evidence of predictive inference generation was only found in the judgment task, but not in the lexical decision task, under dual-task conditions.

3.2.5 Conclusion of Experiment 2

The goal of Experiment 2 was to gain insight into the relationship between demands on cognitive resources and generation of predictive inferences during EFL reading by employing a dual-task paradigm. The results of lexical decision times confirmed and extended the findings of Experiment 1. First, the results under a single-task condition partially replicated the results of Experiment 1, providing further evidence for the view that Japanese EFL readers are more likely to make predictive inferences when the inferences are related to narrative characters' motivation and are strongly constrained by the context (support for H1). Second, the results under dual-task conditions found some impairment in predictive inference generation when fewer cognitive resources are available during reading, although the effect was not strong (the answer to RQ2). Although previous L2 studies have indicated that there is a relationship between inference generation and the amount of cognitive resources available during reading, these studies have focused on the generation of several types of inferences (Horiba, 1996) or on broad categories of inferences (Yoshida, 2003), and were evaluated using the think-aloud method. In contrast, the present experiment examined the generation of a specific type of inference (predictive inferences) using a lexical decision task under dual-task conditions. Nevertheless, the results of lexical decision partially support the view of these previous L2 studies: Increased demands on cognitive resources during reading had some inhibition effects on the generation of predictive inferences.

Furthermore, combining the results of the lexical decision and probability judgment tasks raises the possibility that demands on cognitive resources might impair specific aspects of inference generation. That is, rather than impair generation itself, it is possible that increased demands on cognitive resources during reading decrease the automaticity of predictive inference generation or extend the time required to generate these inferences. This is an interesting implication of this experiment, but further investigations are needed to confirm this possibility and specify the mechanism of inference generation under a resource-limited condition.

Experiment 2 examined impairment of predictive inference generation among Japanese EFL readers, focusing on the amount of cognitive resources available during reading. However, it is also important to clarify how predictive inference generation can be facilitated during EFL reading. Thus, Experiment 3 explored the facilitation effects of strategy and task instructions on predictive inference generation among Japanese EFL readers.

3.3 Experiment 3: Predictive Inference Generation and Strategic Processing in EFL Reading

3.3.1 Purpose, Hypothesis, and Research Questions

In Experiment 1, predictive inferences were not made in Japanese EFL learners' reading when these inferences were less related to narrative characters' goals or motivation (i.e., consequence inferences), even though the context strongly constrained the possible inferences. Nevertheless, it should be noted again that participants in Experiment 1 were instructed simply to read the passages for accurate comprehension. As mentioned in 2.3.2.1, reading goals or task instructions can be an influential factor in predictive inference generation (Allbrriton, 2004; Magliano et al., 1999; van den Broek et al., 2001). Therefore, it is possible that Japanese EFL readers are more likely to make predictive inferences when they are encouraged to process the text strategically (i.e., explicitly instructed to predict what will happen next).

The primary purpose of this experiment, Experiment 3, was to determine whether predictive inferences are strategically made during EFL reading following instructions aimed at active predicting. Of the two subcategories of predictive inferences, consequence inferences, rather than motivational inferences, were tested. To clearly examine the effects of strategy instructions on inference processing, this experiment focused on inferences that Japanese EFL readers are less likely to draw during normal reading.

The main variable in this experiment was orienting instructions: Participants in the *non-orienting condition* were asked simply to understand the passages as in Experiment 1, whereas those in the *strategic orienting condition* were instructed to actively anticipate likely outcomes of the described events in each passage. As noted by Horiba (2013), L2 readers may alter their text processing more distinctively depending on instruction type if they are asked to perform an additional behavioral task. Therefore, to confirm that participants in the strategic orienting condition engaged in intended text processing, they were also asked to write a sentence that would continue the story after reading each passage (Allbritton, 2004).

Some L2 studies have also suggested that the effects of strategy or task instructions are not clearly observed in reading outcomes (e.g., text comprehension assessed by written recall tasks; Horiba, 2013; Yoshida, 2012). Similarly, previous L1 studies have found that strategy instructions for predictive inferences do not affect comprehension of explicit text information, as assessed by written recall (Magliano et al., 1999) or yes/no comprehension tests (Calvo et al., 2006). However, few studies have examined the effect of strategy instructions for predictive inferences on L2 readers' explicit text comprehension. Therefore, this is another area that requires exploration.

Given the previous findings in Experiment 1, it is hypothesized that there would be no evidence of on-line inference generation in the non-orienting condition. The primary focus of this study is whether predictive inferences would be generated on-line in the strategic orienting condition among Japanese EFL readers. The second focus is the effect of strategic processing on Japanese EFL readers' comprehension of explicit text information after reading. Thus, one H and two RQs were addressed in this experiment as follows:

H2: Japanese EFL readers do not automatically make predictive (consequence) inferences when instructed simply to read a passage for comprehension.

RQ3-1: Do Japanese EFL readers strategically make predictive inferences when instructed to anticipate likely outcomes of described events?

RQ3-2: Does strategic processing aimed at predictive inferences affect Japanese EFL readers' comprehension of explicit text information after reading?

Although previous L1 studies have revealed that predictive inference generation is facilitated by explicit instructions to predict what would happen next (Allbritton, 2004; Magliano et al., 1999), the relationship between inference generation and these instructions does not seem to be straightforward in L2 reading. Based on Horiba's (2000, 2013) findings, it is possible that the effect of strategy instructions on predictive inference generation during L2 reading will be small. Alternatively, the magnitude of the effect may differ according to readers' L2 proficiency level, such that high-proficiency readers will be more affected by strategy instructions. Thus, this experiment included readers' L2 proficiency level as a critical reader-related variable.

In this experiment, inference generation during reading was primarily assessed with a lexical decision task on inference-related probe words as in Experiment 2. Furthermore, sentence reading times were also analyzed to confirm the effects of strategy instructions on on-line text processing. Explicit text comprehension after reading was measured with a written recall task.

3.3.2 Method

3.3.2.1 Participants

The participants were 40 Japanese undergraduate and graduate students (21 female and 19 male; aged 18–28 years, M = 20.55, SD = 2.00). All participants had studied English as a foreign language for more than six years as part of formal Japanese education, and they were assumed to have intermediate-level English proficiency.

Two participants' data were removed from the data set due to failure to follow instructions. Additionally, data from one participant were eliminated because the inference activation scores showed extreme outliers in the box plot. No participants made more than 28% errors on the comprehension questions. Therefore, the analyses were based on data from 37 participants.

3.3.2.2 Materials

Passages, target words, and comprehension questions

Thirty-two short narrative texts were used in this experiment. Most of them were chosen from the experimental passages used in Experiment 1, but one passage was adapted from Klin, Guzmán, et al.'s (1999) study specifically for this experiment.[13] These included 16 inference texts and 16 neutral texts (see Appendices 2 and 4).

The inference passages were the CSC version, in which the target inferences were strongly constrained by the context but less related to characters' goals or motivation. The neutral texts were similar to those in Experiment 1. Table 3.26 shows descriptive statistics for the experimental passages used in Experiment 3. Twelve filler texts were also used; these texts were similar in length to the experimental texts, but did not elicit any specific inferences. The experimental texts and orienting conditions were counterbalanced across four material sets using a Latin square. Each set included 16 experimental and 12 filler texts. This ensured that each participant read an equal number of experimental passages in each text type and condition, and that no participant received the same passage twice.

Table 3.26

Number of Words and the Readability of the Experimental Passages in Experiment 3

	M	*SD*	*Max*	*Min*	FKGL	FRE
Inference	50.94	9.16	67	32	5.38	79.68
Neutral	41.19	6.02	53	31	4.73	79.11

Note. FKGL = mean Flesch-Kincaid Grade Level; FRE = mean Flesch Reading Ease.

Each inference and neutral text was paired with a corresponding target word for the

[13] One inference passage whose target word had low familiarity and frequency was replaced with a newly adapted passage.

lexical decision task. Target words were action verbs that represented the predictive inference concept suggested by the inference text (e.g., *break*). As in Experiment 2, only highly frequent and familiar words were used. All target words for experimental passages were three to six letters in length, appeared in the most frequent 2,000-word level (Levels 1 and 2) in the *JACET List of 8000 Basic Words* (JACET Committee of Basic Words Revision, 2003), and had high familiarity ratings on Yokokawa's (2006) 7-point scale. Table 3.27 shows descriptive statistics for the target words used in Experiment 3 (see Appendix 8 for all the target words with their profiles). These factors (i.e., word length, frequency and familiarity) were counterbalanced across four material sets, and there were no significant differences between sets (all *F*s < 1). All target words paired with the filler texts were pseudowords so that the responses in the lexical decision task were balanced.

Table 3.27

Length, Frequency, and Familiarity of Target Words in Experiment 3

	M	*SD*	*Max*	*Min*
Length (number of letters)	4.31	0.79	6	3
Frequency (JACET 8000 Level)	1.38	0.50	2	1
Familiarity (on a 7-point scale from Yokokawa, 2006)	5.49	0.71	6.36	3.96

As in previous experiments, a simple yes/no comprehension question was used for each passage to ensure that participants engaged in text reading. These questions concerned an explicitly stated piece of information.

L2 reading proficiency test

An L2 reading proficiency test was created based on the reading subsection of The

EIKEN Test in Practical English Proficiency (Eiken Foundation of Japan). The EIKEN test has been widely used to measure Japanese EFL learners' English proficiency. Several passages and questions were chosen from retired copies of Grades Pre-1, 2, and Pre-2 of the EIKEN tests (Obunsha, 2009, 2010a, 2010b) as participants' English proficiency was assumed to be approximately intermediate-level.[14] The proficiency test included five passages, each of which was paired with three to five multiple-choice questions, for a total of 20 items.[15]

3.3.2.3 Procedure

In the experiment, participants were tested individually and sessions lasted approximately 75 minutes. The English proficiency test (30 minutes) was completed before the main experimental session.

The main session of the experiment was computer-based and administered using SuperLab 4.5 software (Cedrus, U.S.). Participants were randomly assigned to one of four material sets. First, they read general instructions about the experiment on the computer screen. The instructions informed participants that the session consisted of two phases, and that in each phase they would read short narrative passages in English and make a series of yes/no responses to target words. They were also informed that at the end of each phase, they would be asked to recall the content of some of the passages.[16] Participants completed three practice items before each phase following the same procedure used in the experimental session.

Phase 1 was conducted following the general instructions. At the beginning of this

[14] According to Dunlea (2010), Grades Pre-2 to Pre-1 of the EIKEN Test correspond to the A-2 to B-2 levels of the Council of Europe's (2001) *Common European Framework of Reference for Languages*.
[15] The passages and questions used in the proficiency test are not presented in this dissertation due to copyright restrictions.
[16] Participants were informed of the later recall task to avoid possible learning effects from the first phase on recall performance in the second phase.

phase, participants were instructed to read passages so that they could correctly answer yes/no comprehension questions presented after each passage (i.e., non-orienting condition). Then participants followed the procedure shown in Figure 3.6. In each trial, participants read the text sentence-by-sentence and the reading times for each sentence were recorded in milliseconds by SuperLab. Next, as in previous experiments, participants performed the lexical decision task for the target word and answered a comprehension question. This trial sequence was repeated for each of the 14 passages (eight experimental and six filler passages). After reading all the passages in Phase 1, participants engaged in a cued recall task for the inference texts they had read. Participants received a booklet that displayed the first sentences of these texts. Using the given sentences as cues, they were then asked to write down as much about the passages as possible in Japanese. Participants were allowed an unlimited amount of time to complete the recall task. After the recall task, a short break was provided before starting Phase 2.

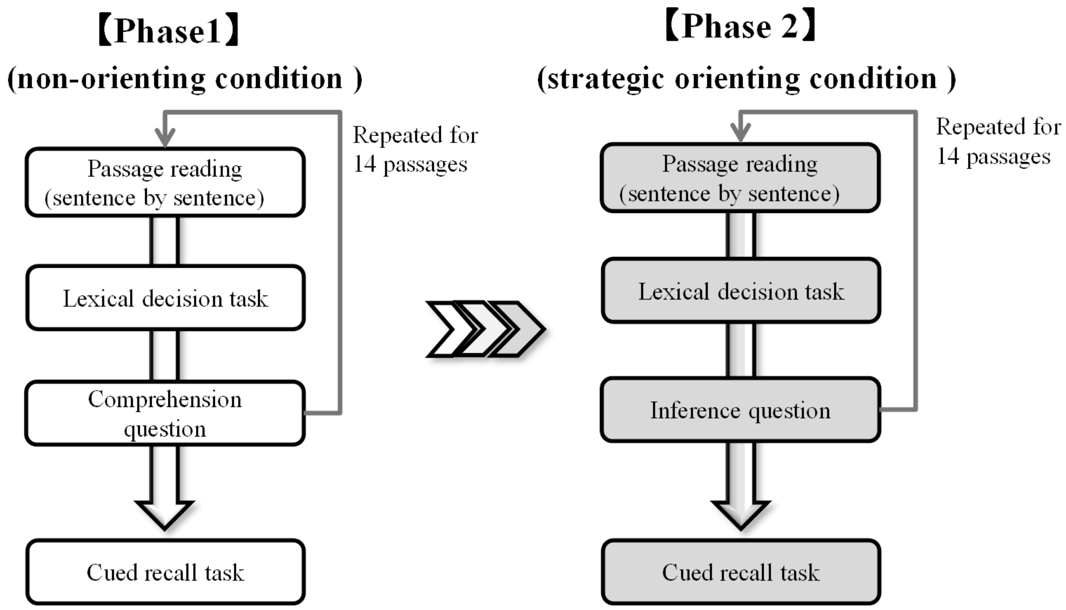

Figure 3.6. Overview of the procedure of the two phases in Experiment 3.

The procedure in Phase 2 was the same as that in Phase 1, with the following exceptions. First, at the beginning of Phase 2, participants were instructed not only to read each sentence for comprehension, but also to predict what would happen next in the story (i.e., strategic orienting condition). More specifically, they were instructed to "think about the likely consequences of the events, actions, and states" in the situations described in each passage (Calvo et al., 2006, p. 71; Magliano et al., 1999). Second, instead of a comprehension question (as in Phase 1), an inference question "What will happen next?" appeared on the screen following the lexical decision task (Allbrriton, 2004) in Phase 2. Participants were asked to write down their answers to the questions in Japanese on a sheet of paper. Similar to Phase 1, after reading all the passages in Phase 2, a cued recall task was completed.

In this experiment, Phase 2 always occurred after Phase 1; that is, the strategic orienting condition always followed the non-orienting condition. The purpose of this order manipulation was to avoid learning effects from the strategic orienting condition on the non-orienting condition. In other words, this manipulation prevented the carryover of strategic reading from the previous trials.

3.3.2.4 Scoring and Analysis

Response data for the lexical decision task as well as comprehension questions were eliminated if participants mistakenly skipped a sentence in the passage. In addition, lexical decision times falling more than 2.5 standard deviations above the mean for each participant were excluded. This resulted in the exclusion of 5.24% of the lexical decision dataset. As in Experiment 2, inference activation scores were first calculated by subtracting a participant's mean correct response time for inference texts from the mean time for neutral texts in each orienting condition. Overall, mean activation scores were then calculated for each proficiency group (higher and lower) and each orienting condition (non-orienting and strategic orienting).

If inference generation is facilitated by strategy instructions, activation scores should be higher in the strategic orienting condition than in the non-orienting condition.

To better understand the effects of strategy instructions on on-line text processing, mean reading times were calculated for each of the four text sentences. Because each text and sentence had a different number of syllables and words (e.g., inference texts contained an average of 13.13 more syllables than neutral texts), sentence reading times were divided by the number of syllables included in each sentence. As in the analysis of lexical decision times, reading times that were 2.5 standard deviations beyond each participant's mean were removed (2.96% of the dataset).

To score recall performance, two raters parsed each inference text into idea units (IUs) by following a procedure used in prior studies (Carrell, 1992; Ikeno, 1996). Each IU consisted of a single clause (main or subordinate, including adverbial and relative clauses). Inter-rater reliability was high for IU segmentation (r = .95) and any disagreements were resolved through discussion. The two raters then scored 30% of recall protocols separately, providing one point if participants either reproduced an IU verbatim or paraphrased it. This process resulted in 91.60% agreement. After resolving disagreements, the remaining data were scored by one rater. All recall scores were transformed into a percentage of the total number of IUs in each passage (except for the first sentence used as a retrieval cue). An arcsine transformation was also performed on the recall production rates before analysis because each text had a different number of IUs, ranging from five to 11.

3.3.3 Results

3.3.3.1 L2 Reading Proficiency Test

After excluding seven low-discriminability items, the reliability of the L2 reading proficiency test was acceptable (Cronbach's α = .70). Before the main analysis, participants were classified into two proficiency groups: the higher (n = 16) and the lower (n = 21) according to a median split of test scores. There was a significant difference in test scores between these two groups, $t(35) = 6.67$, $p < .001$, $d = 2.21$. Table 3.28 shows the proficiency test scores for each proficiency group.

Table 3.28

Mean Scores on the L2 Reading Proficiency Test in Experiment 3

	M	*SD*	*Max*	*Min*
Higher (n = 16)	11.50	0.63	13.00	11.00
Lower (n = 21)	7.71	2.19	10.00	4.00

Note. Maximum possible score is 13.

3.3.3.2 Comprehension Questions

Overall mean correct response rates for the comprehension questions in the non-orienting condition were quite high and above 90% (M = 93.28, SD = 8.22). This supports the claim that participants engaged in text reading even in the absence of specific strategy instructions. Table 3.29 shows the mean accuracy (%) on comprehension questions for each text type and proficiency group.

A 2 (Text Type: inference, neutral) × 2 (Proficiency: higher, lower) ANOVA was conducted on mean correct response rates with Text Type as a within-participants variable and with Proficiency as a between-participants variable. The results revealed no significant main effects of either Text Type or Proficiency. There was also no significant interaction between Text Type and Proficiency (see Table 3.30).

Table 3.29

Mean Accuracy (%) on the Comprehension Questions for Experimental Passages in Experiment 3

	Inference		Neutral	
Proficiency	*M*	*SD*	*M*	*SD*
Higher	95.31	10.08	95.31	13.60
Lower	93.67	11.72	91.67	12.08
Total	94.38	10.92	93.24	12.70

Table 3.30

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Mean Comprehension Question Accuracy in Experiment 3

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency	0.01	1	0.01	0.93	.341	.03
Error (P)	0.48	35	0.01			
			Within-participants			
Text Type (T)	0.00	1	0.00	0.12	.728	.00
T × P	0.00	1	0.00	0.12	.728	.00
Error (T)	0.52	35	0.02			
Total	1.01	73				

3.3.3.3 Lexical Decision Times and Accuracy

Table 3.31 shows mean correct response times for the lexical decision task in each condition. Figure 3.7 displays inference activation scores as a function of orienting condition and readers' L2 proficiency. The error rates were quite low in each condition, and overall mean response accuracy was 96.45%. This high success rate suggests that the target words used in Experiment 3 were easily accessed and processed by the participants, as in Experiment 2.

A 2 (Orienting Condition: non-orienting, strategic orienting) × 2 (Proficiency: higher, lower) ANOVA was conducted on mean activation scores with Orienting Condition as a within-participants variable and with Proficiency as a between-participants variable. The ANOVA results revealed significant main effects of Orienting Condition and L2 proficiency. Specifically, activation scores were higher under the strategic orienting condition than under the non-orienting condition (non-orienting: M = 8.62, SD = 91.33; strategic orienting: M = 53.68, SD = 119.70). Furthermore, scores were higher for readers with higher compared with lower L2 proficiency (lower: M = 4.09, SD = 16.75; higher: M = 66.67, SD = 19.19).

Table 3.31

Mean Correct Lexical Decision Times (ms) of Target Words in Experiment 3

	Non-orienting reading				Strategic orienting reading			
	Inference		Neutral		Inference		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	873	253	892	272	935	288	1,050	345
Lower	874	205	875	172	910	240	917	241
Total	873	224	882	217	921	259	975	294

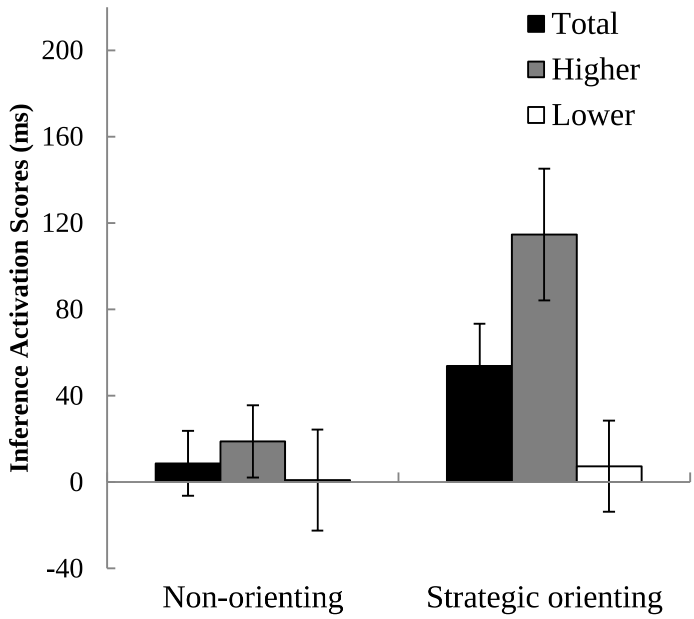

Figure 3.7. Mean inference activation scores (ms) as a function of orienting condition and L2 proficiency. Error bars represent the standard errors of the means. Higher × Non-orienting: M = 18.77, SD = 66.96; Higher × Strategic orienting: M = 114.58, SD = 122.08; Lower × Non-orienting: M = 0.89, SD = 107.26; Lower × Strategic orienting: M = 7.29, SD = 96.88.

More importantly, there was a significant interaction effect between Orienting Condition and Proficiency (see Table 3.32). To examine the simple main effects of Orienting Condition, the means for non-orienting and strategic orienting conditions were compared for each proficiency group using the Bonferroni correction (i.e., adjusted p value < .025). The simple main effect was significant for higher proficiency readers (non-orienting: M = 18.77, SD = 66.96; strategic orienting: M = 114.58, SD = 122.08), $t(15) = 3.50$, $p = .003$, $d = 0.97$, such that there were higher activation scores in the strategic orienting than in the non-orienting condition. This effect was not significant for lower proficiency readers (non-orienting: M = 0.89, SD = 107.26; strategic orienting: M = 7.29, SD = 96.88), $t(20)$ = 0.20, p = .840, d = 0.06. Similarly, mean activation scores for higher and lower proficiency readers were compared for each orienting condition. The simple main effect of proficiency was not significant in the non-orienting condition, $t(35) = 0.56$, $p = .563$, $d = 0.16$, but was significant in the strategic orienting condition, $t(35) = 2.98$, $p = .005$, $d = 0.99$, with higher activation scores for higher proficiency than lower proficiency readers.

Table 3.32

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Orienting Condition Inference Activation Scores in Experiment 3

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency	71139.25	1	71139.25	6.04	.019	.15
Error (P)	412342.00	35	11781.20			
			Within-participants			
Orienting Condition (O)	47424.39	1	47424.39	5.60	.024	.14
P × O	36300.51	1	36300.51	4.29	.046	.11
Error (O)	296279.58	35	8465.13			
Total	863485.73	73				

One-sample *t* tests were also conducted on mean inference activation scores for each condition and proficiency group. The results revealed that activation scores for higher proficiency readers were significantly greater than zero in the strategic orienting condition, and the effect size was large, $t(15) = 3.75$, $p = .002$, $d = 0.91$. However, the scores in all the other conditions were not significantly greater than zero and the effect sizes were comparatively small, non-orienting × higher: $t(15) = 1.12$, $p = .280$, $d = 0.27$; non-orienting × lower: $t(20) = 0.03$, $p = .976$, $d = 0.01$; strategic orienting × lower: $t(20) = 0.35$, $p = .734$, $d = 0.07$.

Table 3.33 shows mean response accuracy of the lexical decision task in each condition.

Facilitation of response accuracy was analyzed to examine the possibility of a speed–accuracy trade-off. A 2 (Orienting Condition: non-orienting, strategic orienting) × 2 (Proficiency: higher, lower) ANOVA was conducted on facilitation of response accuracy (i.e., differences in response accuracy between inference and neutral texts) (non-orienting × higher: M = –1.56, SD = 11.06; non-orienting × lower: M = 2.38, SD = 15.62; strategic orienting × higher: M = 3.13, SD = 8.54; strategic orienting × lower: M = 0.00, SD = 7.91). The results showed no main effects of Orienting Condition or Proficiency (see Table 3.34). The interaction between Orienting Condition and Proficiency was also not significant. It should be noted again that response error rates were quite low in each condition; thus, there was no evidence of a speed–accuracy trade-off.

Table 3.33

Mean Lexical Decision Accuracy (%) of Target Words in Experiment 3

	Non-orienting reading				Strategic orienting reading			
	Inference		Neutral		Inference		Neutral	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	95.31	10.91	96.88	12.50	100.00	0.00	96.88	5.46
Lower	94.05	10.08	91.67	14.43	98.81	5.46	98.81	8.54
Total	94.59	10.43	93.92	13.70	99.32	4.11	97.97	6.92

Table 3.34

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Orienting Condition on Facilitation of Lexical Decision Accuracy in Experiment 3

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participant			
Proficiency (P)	0.00	1	0.00	0.02	.880	.00
Error (P)	0.46	35	0.01			
			Within-participants			
Orienting Condition (O)	0.00	1	0.00	0.19	.666	.01
O × P	0.02	1	0.02	1.78	.190	.05
Error (O)	0.45	35	0.01			
Total	0.93	73				

3.3.3.4 Sentence Reading Times

To better understand the effects of strategy instructions on on-line text processing, mean reading times per syllable (see Table 3.35) were analyzed in a mixed three-way multivariate analysis of variance (MANOVA). The dependent variables were mean reading times for first to fourth sentences, and the independent variables were Text Type, Orienting Condition, and L2 proficiency. The results indicated significant multivariate main effects of Text Type and Orienting Condition (see Table 3.36). Any other potential main or interaction effects were neither significant nor marginally significant.

Table 3.35

Mean Sentence Reading Times (ms) per Syllable in Experiment 3

	Non-orienting reading				Strategic orienting reading			
	Inference		Neutral		Inference		Neutral	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
First								
Higher	449	186	493	201	462	146	486	160
Lower	537	158	546	231	495	171	590	202
Total	499	174	523	217	481	160	545	190
Second								
Higher	405	124	459	133	376	123	455	143
Lower	477	143	503	162	503	206	536	222
Total	446	138	484	150	448	184	501	193
Third								
Higher	385	88	414	161	381	115	467	126
Lower	442	122	509	197	476	184	563	146
Total	417	111	468	186	435	163	522	144
Forth								
Higher	373	108	368	146	431	140	436	142
Lower	460	181	468	147	560	228	594	214
Total	422	158	425	153	504	203	526	200

Table 3.36

Summary Table for Three-Way MANOVA of the Effects of Proficiency, Text Type and Orienting Condition on Sentence Reading Times

Source	Pillai's Trace	*F*	*df*	Error *df*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.19	1.88	4	32	.140	.19
			Within-participants			
Text Type (T)	0.52	8.64	4	32	.000	.52
T × P	0.05	0.45	4	32	.771	.05
Orienting Condition (O)	0.42	5.78	4	32	.001	.42
O × P	0.10	0.91	4	32	.468	.10
T × O	0.04	0.35	4	32	.843	.04
T × O × P	0.12	1.04	4	32	.401	.12

Follow-up univariate repeated ANOVAs were conducted on reading times for each sentence to examine on what sentences the effects of Text Type and Orienting Condition were prominent. The results revealed that the main effect of Text Type was significant for all sentences except for the fourth, with shorter reading times for inference texts than neutral texts. More importantly, the main effect of orienting condition was significant only for the third and fourth sentences, but not for the first and second sentence (see Table 3.37). As shown in Table 3.35, both proficiency groups of participants took a longer time reading the third and fourth sentences in the strategic orienting condition than in the non-orienting condition. This trend was consistent between inference and neutral texts.

Table 3.37

Summary Table for Follow-Up One-Way ANOVAs of the Effects of Text Type and Orienting Condition on Sentence Reading Times

Source		*SS*	*df*	Error *df*	*MS*	*F*	*p*	η_p^2
Text Type	First	67899.82	1	35	67899.82	5.40	.026	.13
	Second	83369.38	1	35	83369.38	5.98	.020	.15
	Third	166566.71	1	35	166566.71	30.81	.000	.47
	Forth	4179.37	1	35	4179.37	0.72	.403	.02
Orienting Condition	First	192.17	1	35	192.17	0.02	.900	.00
	Second	1417.56	1	35	1417.56	0.20	.658	.01
	Third	42659.83	1	35	42659.83	8.50	.006	.20
	Forth	281806.72	1	35	281806.72	22.01	.000	.39

3.3.3.5 Cued Recall

The overall reliability of the recall test was high (Cronbach's $\alpha = .86$).[17] Mean recall production rates (%) are shown in Table 3.38. A mixed two-way ANOVA was then conducted on recall performance, with Orienting Condition as a within-participants variable and Proficiency as a between-participants variable. The results revealed a marginally significant main effect of orienting condition, with better performance in the strategic orienting condition than the non-orienting condition. Neither the main effect of proficiency nor the interaction were significant or marginally significant (see Table 3.39).

[17] The reliability for the non-orienting condition ranged from .65 to .80 ($M = .72$, $SD = .08$). The reliability for the strategic orienting condition ranged from .81 to .85 ($M = .83$, $SD = .02$).

Table 3.38

Mean Recall Rates (%) of Explicit Text Information in Experiment 3

	Non-orienting reading		Strategic orienting reading	
Proficiency	*M*	*SD*	*M*	*SD*
Higher	57.27 (48.80)	14.86 (11.09)	59.32 (51.18)	19.68 (14.97)
Lower	47.15 (42.47)	13.39 (11.66)	51.03 (48.48)	21.71 (12.92)
Total	50.92 (45.21)	15.45 (11.70)	55.30 (49.65)	20.79 (13.71)

Note. The values after an arcsine transformation are shown in parentheses.

Table 3.39

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Orienting Condition on Recall Performance in Experiment 3

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	370.08	1	370.08	1.57	.218	.04
Error (P)	8226.70	35	235.05			
			Within-participants			
Orienting Condition (O)	320.20	1	320.20	3.69	.063	.10
O × P	60.08	1	60.08	0.69	.411	.02
Error (O)	3037.21	35	86.78			
Total	12014.28	73				

3.3.4 Discussion

H2: Japanese EFL readers do not automatically make predictive (consequence) inferences when instructed simply to read a passage for comprehension.

The lexical decision time data did not show high inference activation scores in the non-orienting condition. Indeed, the activation scores in the non-orienting condition were not significantly above zero among either higher or lower proficiency readers. These findings support the hypothesis regarding predictive inference generation in the absence of strategy instructions. Specifically, consistent with Experiment 1, the present study confirmed that when given instructions simply to read the passage for comprehension, Japanese EFL readers are less likely to generate consequence inferences on-line even though the predictability of the to-be-inferred event is high.

Mean answer rates for the comprehension questions in the non-orienting condition were above 90%. In addition, there was no significant difference in the answer rates between text types. These results support the claim that participants were able to comprehend the experimental passages, regardless of text types. Consequently, the possibility that participants did not make inferences in the non-orienting condition due to poor comprehension of text meaning was ruled out.

RQ3-1 Do Japanese EFL readers strategically make predictive inferences when instructed to anticipate likely outcomes of described events?

Regarding the strategic generation of predictive inferences, the results pertaining to lexical decision times revealed a significant main effect of orienting condition on inference activation scores, with higher scores in the strategic condition than in the non-orienting condition. This suggests that providing strategy instructions should facilitate predictive inference generation. Analysis of reading time partially supports the claim that inference generation was facilitated in the strategic orienting condition. In particular, results demonstrated that participants had read the latter part of the text (especially the fourth sentence) for a significantly longer amount of time in the strategic-orienting condition than in the non-orienting condition. This finding suggests that participants altered their text processing according to instruction type by allocating increased effort toward making inferences in the latter part of the text.

However, the significant interaction effect between orienting condition and L2 proficiency on inference activation scores revealed that the effect of strategy instructions was more prominent among higher than lower proficiency readers. Indeed, inference activation scores were significantly greater than zero only for higher proficiency readers under the strategic condition. Horiba (2000, 2013) suggested that flexibility in controlling text processing according to task instructions is affected by readers' language proficiency. Consistently, higher proficiency readers in this study would be better than would lower proficiency readers be at controlling text processing according to instructions. In addition, Murray and Burke (2003) found that skilled L1 readers are more likely to generate predictive inferences on-line than are less skilled readers. The present study denotes similar facilitation effects of L2 reading proficiency, although these effects were enhanced by strategic orienting instructions.

As noted above, analysis of reading time suggests that both higher and lower proficiency readers altered their text processing according to instructions. Given this result, it seems more plausible to state that higher and lower proficiency readers might differ in their ability to activate a particular inference concept during reading according to instructions. Thus, less proficiency readers might possibly have failed to make inferences due to the low-level of inferential activation, even though they were aware of the instructions and their processing mode was altered to some extent.

One might wonder why higher proficiency readers tend to take more time to respond to lexical decision probes than less proficient readers do (see Table 3.30). Although there were no significant differences between higher and lower proficiency readers' response times in any of the four conditions (all *t*s < 1.5), response times seem especially slow among higher proficiency readers for the neutral texts in the strategic orienting condition. Given that the neutral texts did not induce specific predictions, this result may be attributed to the inhibition of response times caused by activating inference concepts unrelated to the target probes. Nevertheless, the results consistently suggest that proficient readers should more actively engage in inferential processing during reading than less proficient readers should when they were encouraged to do so.

RQ3-2: Does strategic processing aimed at predictive inferences affect Japanese EFL readers' comprehension of explicit text information after reading?

The results of the written recall task demonstrated that both higher and lower proficiency readers were likely to recall more information from the text in the strategic orienting condition than the non-orienting condition. Therefore, it is possible that strategy instructions about anticipating the outcomes of described events enhance readers' comprehension of explicit text information.

This effect pertaining to strategy instructions was assumed to be the result of readers' active engagement in text processing. In the strategic orienting condition, readers need to process the text more actively and carefully to determine how the text should constrain the possible outcome of events. The aforementioned longer sentence reading time in the strategic orienting condition is indicative of readers' active and careful engagement with the text. Readers also answered a literal yes/no comprehension question in the non-orienting condition, whereas they wrote down a continuing sentence in the strategic orienting condition. This additional task in the strategic orienting condition might further encourage reprocessing or rehearsal of the text. Such careful and repeated processing of the text, therefore, could have strengthened readers' memory for each passage.

In contrast to the results of lexical decision times, there was no significant interaction effect of L2 proficiency and orienting condition on recall performance. Thus, recall performance tended to be better in the strategic orienting condition than in the non-orienting condition, regardless of readers' L2 proficiency level. This is the same direction as the results pertaining to reading times for the third and fourth sentences (i.e., longer reading times in strategic orienting condition than in the non-orienting condition for both proficiency groups of participants), which supports the view that readers' active and careful engagement with the text should enhance comprehension of explicit text information. Therefore, rather than predictive inference generation itself, active processing of the text would rather be required to improve explicit text comprehension.

Calvo et al. (2006) and Magliano et al. (1999) demonstrated that strategic processing for predictive inferences had no impact on L1 readers' explicit text comprehension. Although the present study suggests positive effects of strategic processing on Japanese EFL readers' text comprehension, the results are consistent with these past studies in that strategic processing for predictive inference generation did not reduce explicit text comprehension. One might expect that strategically making predictions during reading is demanding for EFL readers, resulting in impaired comprehension because they require more cognitive effort to comprehend a text than L1 readers did. However, this experiment demonstrates that EFL readers' comprehension of explicit text information is not impaired when their attention is focused on generating predictive inferences during reading. Nevertheless, it should be noted that the experimental passages used in this study had relatively low WM demands. The passages were short and consisted of simple words and sentences. It is possible that these factors enabled strategic text processing with little cost to the extraction of explicit text information.

3.3.5 Conclusion of Experiment 3

The results of Experiment 3 confirm and extend findings in Experiment 1. First, consistent with Experiment 1, lexical decision data demonstrated that predictive inferences (that were not related to characters' goals or motivation) were less likely to be generated on-line during Japanese EFL reading when readers were instructed simply to read the passage for accurate comprehension (a support for H2). Second, lexical decision and sentence reading times presented the facilitation effects of strategy instructions on predictive inference generation during Japanese EFL reading, but the facilitation effect of strategy instructions on inference generation was much stronger among readers with higher L2 proficiency than among those with lower L2 proficiency (answer to RQ3-1). Third, recall production analysis found that strategic processing for predictive inference generation was achieved without reducing L2 readers' comprehension of explicit text information; rather, it likely facilitated comprehension for both higher and lower L2 proficiency readers (answer to RQ3-2). The following section summarizes the main findings of Experiments 1 to 3, and concludes Study 1.

3.4 Conclusion of Study 1

Study 1 examined whether and how Japanese EFL learners make predictive inferences in reading, focusing on the effects of text characteristics (Experiment 1), the amount of cognitive resources available during reading (Experiment 2), and strategy and task instructions (Experiment 3).

Experiment 1 investigated the effects of two text factors on predictive inference generation, subtypes of inferences and contextual constraint. Recognition times for target words and recall productions of inferences showed that predictive inferences were most likely to be generated in reading when the inferences were motivational (i.e., related to motivation for narrative characters' actions) and strongly constrained by context. In contrast, no evidence of predictive inference generation was found with consequence inferences (i.e., less related to narrative characters' motivation), regardless of the strength of contextual constraint.

Experiment 2 examined how the amount of cognitive resources available during reading affects predictive inference generation by employing a dual-task methodology. In a single-task condition, lexical decision times for target words confirmed the finding of Experiment 1: Predictive inferences are likely to be made when motivational inferences are strongly constrained by context. However, in a dual-task condition that required participants to hold word lists in memory while reading, the results of lexical decision times suggested some impairment of predictive inference generation, even when the texts strongly induced the inferences. In addition, the results of probability judgments for target sentences raised the possibility that increased demands on cognitive resources specifically reduced the automaticity and immediacy of predictive inference generation.

Finally, Experiment 3 tested whether the generation of consequence inferences (found less likely to be made during normal reading in Experiment 1) is facilitated by strategy and task instructions, including the consideration of learners' L2 proficiency level. The results of lexical decision times refined the finding of Experiment 1: In the absence of specific strategy instructions, predictive inferences that are not related to narrative characters' motivation are less likely to be made during reading. The results of sentence reading times suggested that both higher and lower proficiency readers altered their text processing to some extent when given strategy instructions aimed at predictive inference generation. However, lexical decision times revealed that the facilitation effects of strategy instructions on inference generation were much larger among higher than lower proficiency readers. Furthermore, recall productions showed that giving instructions that focused attention on implicit predictive information did not impair readers' comprehension of explicit text information.

The results of these experiments suggest that predictive inferences can be generated in Japanese EFL leaners' reading comprehension, and the generation of these inferences is affected by similar factors suggested in previous L1 and L2 reading research (i.e., text characteristics, cognitive demands in reading, and strategic processing). However, it should be noted that predictive inference generation among Japanese EFL readers seems more limited than among L1 English readers. For instance, consequence inferences are less likely to be made during EFL reading without strategy instructions, even though the future narrative event is highly predictable. Furthermore, the facilitation effect of strategy instructions on predictive inference generation seems not as prominent as in L1 reading; these instructions especially facilitate the inference generation among higher L2 proficiency readers.

Study 1 investigated the generation of predictive inferences in Japanese EFL learners' reading comprehension. However, as noted in the introduction, readers do not always make predictive inferences that are consistent with the subsequent context, and they need to revise the drawn inferences when the inferences are later disconfirmed. Thus, Study 2 was conducted to examine the revision of predictive inferences in Japanese EFL learners' reading comprehension.

Chapter 4

Study 2: Revising Predictive Inferences in Japanese EFL Learners' Reading Comprehension

4.1 Experiment 4: Examining the Revision of Predictive Inferences in EFL Reading

4.1.1 Purpose and Research Questions

The primary purpose of Experiment 4 is to examine the revision of initially drawn but later disconfirmed predictive inferences in Japanese EFL learners' reading. To date, several previous studies have been conducted to examine the generation of predictive inferences in reading (see 2.3.2 for a review). However, only a few studies have investigated revision of predictive inferences while reading; unfortunately, their findings were contradictory or unclear (Iseki, 2006; Potts et al., 1988).

As in previous studies (Iseki, 2006; Potts et al., 1988), this experiment utilized two types of inference texts: *predictive* and *disconfirming* texts. First, predictive texts were designed to strongly induce specific predictive inferences during normal reading; the target inference is related to the motivation for a character's action and strongly constrained by the context (i.e., the MSC condition in Experiment 1). Next, disconfirming texts were newly constructed for this experiment by adding a continuing sentence to the predictive texts that disconfirms the target inference suggested by the predictive text. This experiment compared the data obtained from predictive and disconfirming texts.

This experiment investigated the two cognitive mechanisms required for the revision of disconfirmed inferences: (a) suppression, which refers to the decreased activation of inferential information during reading; and (b) deletion, which refers to the elimination of inferential information from long-term text memory (see also 2.3.3). These two mechanisms were examined by individual tasks: Suppression of inferences was assessed based on recognition times for target words, whereas deletion of inferences was investigated by analyzing the intrusion of target inferences in a cued recall task.

In addition to the revision of predictive inferences, Experiment 4 investigated how the disconfirmation of drawn predictive inferences affects comprehension of explicit text information. Fincher-Kiefer (1993) noted that making predictions that are inconsistent with future text is prohibitive because revising text comprehension may be cognitively demanding. Indeed, Nahatame (2010, 2011) suggested the negative effects of making incorrect predictive inferences on memory and processing of explicit text information among less proficient L2 readers. However, because few studies have addressed this issue, more evidence needs to be provided. Thus, the second purpose of this experiment is to examine how L2 learners' explicit text comprehension is affected by the disconfirmation of predictive inferences.

To achieve these goals of this experiment, three RQs were addressed in Experiment 4:

RQ4-1: Do Japanese EFL readers suppress the activation of predictive inferences when the subsequent context disconfirms the inferences?

RQ4-2: Do Japanese EFL readers delete the disconfirmed predictive inferences from long-term text memory?

RQ4-3: How does the disconfirmation of predictive inferences influence Japanese EFL readers' comprehension of explicit text information?

As noted in 2.1.1, the SBF supposes that skilled readers are characterized as having an effective suppression mechanism (Gernsbacher, 1990; Gernsbacher et al., 2004). Similarly, Ushiro (2010) found that skilled L2 readers revise text comprehension more successfully than less skilled L2 readers. Therefore, this experiment investigated the revision of predictive inferences including consideration of learners' L2 reading proficiency level.

4.1.2. Method

4.1.2.1 Participants

The participants of this experiment were 37 students (19 male and 18 female; aged 18–21 years, M = 19.11, SD = 1.13) from the same Japanese university. They were in different years (first o fourth) and had different majors (e.g., humanities, international studies etc.). All participants had studied English as a foreign language for more than six years as part of formal Japanese education, and they were assumed to have intermediate-level English proficiency.

All data from one participant with more than 28% errors in the comprehension questions was excluded. Additionally, data from one participant were eliminated because the inference activation scores showed extreme outliers in the box plot. Therefore, the analyses were based on data from 35 participants.

4.1.2.2 Materials

Passages, target words, and comprehension questions

Thirty-two short narrative texts were adopted from Experiment 1. These included 16 inference (i.e., predictive) texts and 16 neutral texts. The MSC texts from Experiment 1 were used as inference passages to maximize the likelihood of target inference generation. To examine the revision of inferences, it is necessary to first confirm that participants make target inferences during reading. The neutral texts were same as those used in Experiment 1.

Because the primary purpose of this experiment was to examine the revision of the drawn predictive inferences, another version of the inference texts (i.e., disconfirming texts) was newly constructed. In the disconfirming texts, a single sentence was added after the final sentence of each predictive text. This extra sentence provided the information precluding the inference suggested by the predictive text (see Table 4.1 for an example). The disconfirming sentences for each passage were created by the author, referring to the materials in previous studies (Iseki, 2006; Potts et al., 1988).

Table 4.1

Sample of Experimental Passages, Target Words, and Comprehension Questions in Experiment 4

Inference text (predictive and disconfirming)
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.*
Target word: *hit*
Question: Was the first batter a new boy on the team? (Yes)
Neutral text
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.
Target word: *hit*
Question: Had the young boys disappeared the day before? (Yes)

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

Considering the importance of the disconfirming sentences in this experiment, these sentences were assessed by two other raters using a 7-point scale in terms of their plausibility as sentences following the original text (i.e., predictive text). Subsequently, the sentences rated as less plausible (i.e., less than 4 on a 7-point scale) were modified for increased plausibility through discussion with raters. In addition, all disconfirming sentences were composed of between 15 and 20 simple words. These processes ensured that the different features of disconfirming sentences such as the plausibility, length, and difficulty did not affect the results. All disconfirming sentences had been checked and corrected by a native English speaker. Table 4.2 shows the number of words, lexical frequency of disconfirming sentences, and semantic relatedness to the fourth sentences. These factors were counterbalanced across four material sets, and there were no significant differences between sets (all *F*s < 2.33). All disconfirming texts used in Experiment 4 are shown in Appendix 1.

Table 4.2

Number of Words, Lexical Frequency, and Semantic Relatedness for the Disconfirming Sentences

	M	*SD*	*Max*	*Min*
Number of words	17.00	1.51	20	15
Lexical frequency	1.47	0.45	2.73	1.00
Semantic relatedness to fourth sentences	0.42	0.10	0.83	0.10

Note. Lexical frequency is the mean *JACET 8000* level of words included in the sentence. Semantic relatedness was calculated using Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998; see also Nahatame, 2012) and ranged from 0 to 1.

In sum, two sets of 16 inference passages (i.e., predictive and disconfirming texts) and 16 neutral passages were used in this experiment. Table 4.3 shows descriptive statistics for the experimental passages. In addition, 12 filler texts were also used and these filler texts did not suggest any specific inferences. Some filler texts consisted of five sentences so that they were similar in length to the disconfirming texts.

Table 4.3

Number of Words and the Readability of the Experimental Passages in Experiment 4

	M	*SD*	*Max*	*Min*	FKGL	FRE
Predictive text	49.00	7.54	61	36	4.98	80.81
Disconfirming text	65.13	7.73	77	51	5.01	82.26
Neutral text	41.94	7.87	62	30	3.98	84.44

Note. FKGL = mean Flesch-Kincaid Grade Level; FRE = mean Flesh Reading Ease.

Four material sets were created to counterbalance the passages across versions, and each set included 12 of 16 experimental passages. This confirmed that each experimental passage appeared an equal number of times in the three versions (i.e., predictive, disconfirming, and neutral) and no participant received the same passage twice. All of the filler passages were included in each set.

As in previous experiments, each text had a corresponding target word and a yes/no comprehension question. Target words (also the same as those in Experiment 1) were the verbs that represented the inferential information suggested by the predictive texts, but were precluded by the disconfirming texts. Target words paired with the filler texts were simple words extracted from the passages, including verbs, nouns, or adjectives, to balance the number of yes and no responses of the recognition task. The comprehension questions focused on details of the passage that were not directly relevant to the targeted inference.

L2 reading proficiency test

A 54-item random cloze test was created and used to measure participants' L2 reading proficiency. The cloze test was used as the L2 reading proficiency test in this experiment because of the limitation of experiment time. Specifically, this test includes a high number of items but can be conducted in a short amount of time.

The test was created based on a 417-word text from *Timed Readings*, *Book one* (Spargo, 1989). The number of sentences and words as well as the readability of the text is reported in Table 4.4. This test was constructed by replacing every seventh word in the text with a blank except the first, second, and last sentences in the text. The word next to the seventh word was deleted when the seventh word was a proper noun such as a person's name.[18]

Table 4.4

Number of Sentences, Words and the Readability of the Text Used for the Cloze Test in Experiment 4

Number of sentences	Number of words	FKGL	FRE
36	417	2.9	94.1

Note. FKGL = mean Flesch-Kincaid Grade Level; FRE = mean Flesch Reading Ease.

4.1.2.3 Procedure

As in previous experiments, participants were individually tested in the experiment. The L2 proficiency test (17 minutes) was completed prior to the main experimental session. During the test, participants were asked to write an English word that was contextually appropriate for each blank in the text.

In the main experimental session, participants followed the same procedures as in Experiment 1. Participants were randomly assigned to one of four material sets. They read a passage sentence-by-sentence on the computer screen, performed the recognition task for a

[18] Similar to Experiment 3, the passage used for the cloze test is not presented in this dissertation due to copyright restrictions.

target word immediately after the final sentence of each passage, and then answered a comprehension question. This trial was repeated for each of the 24 passages (12 experimental and 12 filler passages). After completing all trials, participants were provided with 10 minutes to read another set of eight inference passages on paper. They then engaged in a cued recall task, in which the first sentence of each passage was provided as a retrieval cue to recall the subsequent text contents. Participants were given an unlimited amount of time to complete the recall task.

4.1.2.4 Scoring and Analysis

The cloze test was marked using the semantically-acceptable word scoring method. One point was given for each answer where participants produced a word semantically acceptable in the context, even if not an exact word. Thirty percent of the data were randomly selected from the pool of the data and scored by the two raters separately, resulting in an agreement ratio of 91.36%; disagreements were resolved through discussion among the raters. The remaining data were scored by one of the raters.

Regarding the recognition data, participants' data from a passage were eliminated when they mistakenly skipped a sentence of the passage. In addition, recognition times which were three standard deviations beyond the mean for a participant were excluded, along with those less than 100 ms. These processes led to the removal of 1.41% of the data. As in Experiment 1, inference activation scores were first calculated by subtracting a participant's mean correct recognition time for neutral texts from the mean time for inference (i.e., predictive and disconfirming) texts. Overall, mean activation scores were then calculated for each inference condition and proficiency group. If readers activate target inferences during reading, correct responses to target words should be delayed by the activation, resulting in activation scores greater than zero. The present experiment compared the activation scores from the predictive text condition to those from the disconfirming text condition. If the disconfirming context serves to decrease the activation of the inference, then the activation scores in the disconfirming text condition should have been lower than those in the predictive text condition (see Iseki, 2006; Potts et al., 1988; 2.3.3, p. 53).

Recall protocols were scored in terms of two types of information: inferential information and explicit text information. Inferential information was the concepts represented by the target words (e.g., *hit*), as mentioned in Experiment 1. When the target word was included in participants' recall protocols, it was scored as the production of inferential information. If the drawn inference is deleted from the text memory, there should be fewer inference intrusions in the disconfirming text condition than in the predictive text condition. In contrast, explicit text information refers to literal information explicitly described in the passages; that is, the three common sentences included in both the predictive and the disconfirming texts (except for the first sentence that was used as a cue). For the scoring of explicit text information, text sentences were parsed into a set of IUs by the two raters based on a procedure in prior studies (Carrell, 1992; Ikeno, 1996), resulting in high inter-rater reliability, r = .95; any disagreements were resolved through discussion.

Thirty percent of recall protocols were marked by the two raters separately in terms of explicit text and inferential information, resulting in an agreement ratio of 92.09%. After resolving disagreements, the remaining data were scored by one rater (i.e., the author). All recall scores of inferential information were transformed into a proportion, dividing the scores by the total number of passages in each condition. Similarly, all recall scores of explicit text information were transformed into a percentage of the total number of IUs in the common three sentences. Furthermore, an arcsine transformation was performed on the recall production rates of explicit text information because each text had a different number of IUs (five to 11 IUs).

4.1.3 Results

4.1.3.1 L2 Reading Proficiency Test

The L2 proficiency test indicated high reliability (Cronbach's $\alpha = .80$) after excluding 12 low-discriminability items. Prior to the main analysis, participants were divided into higher ($n = 17$) and lower ($n = 18$) proficiency groups, according to a median split of test scores. There was a significant difference in test scores between these groups, $t(33) = 9.37$, $p < .001$, $d = 3.17$. Table 4.5 shows the proficiency test scores for each proficiency group.

Table 4.5

Mean Scores on the L2 Reading Proficiency Test in Experiment 4

	M	*SD*	*Max*	*Min*
Higher (n = 17)	27.47	1.87	32	25
Lower (n = 18)	18.06	3.72	24	11

Note. Maximum possible score is 42.

4.1.3.2 Comprehension Questions

Mean correct answer rates for the comprehension questions were quite high (above 90%) as in previous experiments. Table 4.6 shows the mean correct answer rates of comprehension questions for each text type and proficiency group.

A 3 (Text Type: predictive, disconfirming, neutral) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on mean correct answer rates with Text Type as a within-participants variable and with Proficiency as a between-participants variable. The results found a significant main effect of Proficiency, with higher scores for higher proficiency than lower proficiency readers. Neither a significant main effect of Text Type nor an interaction between Text Type and Proficiency was significant (see Table 4.7).

Table 4.6

Mean Accuracy (%) on the Comprehension Questions for Experimental Passages in Experiment 4

	Predictive		Disconfirming		Neutral	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	95.10	11.05	98.53	6.06	91.18	15.16
Lower	88.89	17.62	86.11	15.39	90.28	15.19
Total	91.90	14.92	91.67	13.36	90.71	14.96

Table 4.7

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Mean Comprehension Question Accuracy in Experiment 4

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.11	1	0.11	6.57	.015	.17
Error (P)	0.56	33	0.02			
			Within-participants			
Text Type (T)	0.00	2	0.00	0.12	.889	.00
T × P	0.06	2	0.03	1.38	.258	.04
Error (T)	1.39	66	0.02			
Total	2.12	104				

4.1.3.3 Recognition Times and Accuracy

The mean recognition times for correct responses as a function of text types and proficiency groups are reported in Table 4.8. Figure 4.1 displays inference activation scores for each text condition and proficiency group.

Table 4.8

Mean Correct Recognition Times (ms) of Target Words in Experiment 4

	Predictive condition		Disconfirming condition		Neutral condition	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	1,168	291	1,288	427	1,042	269
Lower	1,305	398	1,277	367	1,159	287
Total	1,239	352	1,282	391	1,103	281

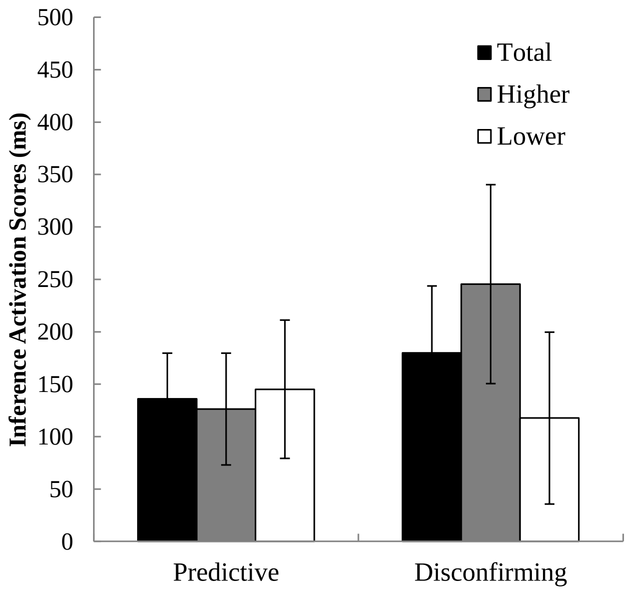

Figure 4.1. Mean inference activation scores (ms) as a function of text type and L2 proficiency in Experiment 4. Error bars represent the standard errors of the means. Higher × Predictive: *M* = 126.23, *SD* = 220.17; Lower × Predictive: *M* = 145.17, *SD* = 294.84; Higher × Disconfirming: *M* = 245.51, *SD* = 391.26; Lower × Disconfirming: *M* = 117.68, *SD* = 366.53.

Prior to the test of the suppression of inferences, it is necessary to confirm that participants actually activated the targeted inferences during reading. Therefore, two-tailed one sample *t* tests were conducted on the inference activation scores in the predictive text condition to determine whether they were significantly above zero. The results showed that the scores were significantly greater than zero for all readers and the higher proficiency group, $t(35) = 3.12$, $p = .004$, $d = 0.52$; $t(16) = 2.36$, $p = .031$, $d = 0.56$, and that the effect approached significance for the lower proficiency group, $t(17) = 2.08$, $p = 0.52$, $d = 0.48$. Given that medium-sized effects were obtained for both groups of participants, it is suggested with high certainty that, for both proficiency groups, targeted inferences were activated during reading.

Then, in order to examine the suppression of activated inferences (RQ4-1), the inference activation scores in the predictive text condition were compared to those in the disconfirming text condition. The mean inference activation scores were submitted to a 2 (Text Type: predictive, disconfirming) × 2 (Proficiency: higher, lower) mixed ANOVA with Text Type as a within-participants variable, and Proficiency as a between-participants variable. The results showed that neither main effects of Text Type nor Proficiency approached significance (see Table 4.9). In addition, an interaction between Text Type and Proficiency was not significant.

Table 4.9

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Inference Activation Scores in Experiment 4

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	51839.57	1	51839.57	0.40	.531	.01
Error (P)	4266525.74	33	129288.66			
			Within-participants			
Text Type (T)	36837.71	1	36837.71	0.45	.509	.01
T × P	94180.34	1	94180.34	1.14	.293	.03
Error (T)	2721144.13	33	82458.91			
Total	7170527.49	69				

Table 4.10 reports the mean accuracy (%) of target word recognition. A 2 (Text Type: predictive, disconfirming) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on accuracy inhibition effects (i.e., differences in response accuracy between inference and neutral texts), with Text Type as a within-participants variable and Proficiency as a between-participants variable. Although accuracy inhibition effects were numerically larger for the disconfirming than predictive text condition (higher × predictive: M = 9.72, SD = 25.92; lower × predictive: M = 7.84, SD = 17.79; higher × disconfirming: M = 14.71, SD = 21.39; lower × disconfirming: M = 13.89, SD = 21.39), a main effect of neither Text Type nor Proficiency was significant (see Table 4.11). Thus, consistent with the results of activation scores, accuracy inhibition effects were similar regardless of text types and L2 proficiency. Nevertheless, the overall mean recognition accuracy was relatively high (M = 88.68%, SD = 10.87). Therefore, in this experiment (as in Experiment 1), only the response time data were

regarded as evidence of inference generation.

Table 4.10

Mean Recognition Accuracy (%) of Target Words in Experiment 4

	Predictive		Disconfirming		Neutral	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	90.69	15.83	83.82	21.54	98.53	6.06
Lower	84.72	21.25	80.56	23.57	94.44	10.69
Total	87.62	18.78	82.14	22.34	96.43	8.88

Table 4.11

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Recognition Accuracy Inhibition in Experiment 4

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.00	1	0.00	0.01	.929	.00
Error (P)	2.03	33	0.06			
			Within-participants			
Text Type (T)	0.05	1	0.05	1.37	.250	.04
T × P	0.00	1	0.00	0.08	.777	.00
Error (T)	1.28	33	0.04			
Total	3.37	69				

4.1.3.4 Cued Recall

Table 4.12 shows the mean proportion of targeted inferences recalled by the higher and lower proficiency groups. Inferential information in the predictive text condition was recalled to the same degree as the data from Experiment 1,[19] suggesting that participants in this experiment maintained target inferences in their long-term text memory when they read the predictive passages.

Table 4.12

Mean Recall Rates (%) of Target Inferences in Experiment 4

	Predictive		Disconfirming	
Proficiency	*M*	*SD*	*M*	*SD*
Higher	10.88	13.95	8.82	12.31
Lower	18.61	22.87	13.89	12.78
Total	14.86	19.19	11.43	12.64

To examine RQ4-2, a 2 (Text Type: predictive, disconfirming) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on the proportion of recalled inferential information, with Text Type as a within-participants variable and Proficiency as a between-participants variable. Similar results were found to those obtained in the analysis of the recognition times. Neither main effects of Text Type nor Proficiency were significant. Additionally, an interaction between Text Type and Proficiency was not significant or marginally significant (see Table 4.13). Thus, the inferential information was recalled in the disconfirming text condition to the same degree as in the predictive text condition, irrespective of readers' L2 proficiency.

[19] A two-tailed paired *t* test indicated that there were no significant differences in the amount of target inference intrusion between Experiments 1 and 4, $t(62) = 1.49$, $p = .143$, $d = 0.37$.

Table 4.13

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Recall Rates of Target Inferences in Experiment 4

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.07	1	0.07	2.76	.106	.08
Error (P)	0.86	33	0.03			
			Within-participants			
Text Type (T)	0.02	1	0.02	0.77	.387	.02
T × P	0.00	1	0.00	0.12	.733	.00
Error (T)	0.86	33	0.03			
Total	1.82	69				

Table 4.14 shows the proportion (%) of recalled explicit text information. Then, to answer RQ4-3, a 2 (Text Type: predictive, disconfirming) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on recall performance, with Text Type (predictive, disconfirming) as a within-participants variable and Proficiency (higher, lower) as a between-participants variable.

The results showed that a main effect of Proficiency was statistically significant (see Table 4.15). Higher proficiency readers significantly recalled more text information than lower proficiency readers. However, neither a main effect of Text Type nor an interaction between Text Type and Proficiency was significant. Although recall performance was numerically poorer for the disconfirming than predictive condition, the amount of explicit information recalled by participants did not significantly differ between the two text conditions. This result was consistent among participants from different proficiency groups.

Table 4.14

Mean Recall Rates (%) of Explicit Text Information in Experiment 4

	Predictive		Disconfirming	
Proficiency	*M*	*SD*	*M*	*SD*
Higher	44.82 (39.82)	19.77 (15.92)	39.29 (35.51)	18.27 (14.89)
Lower	30.44 (28.63)	17.31 (14.54)	26.78 (26.46)	13.18 (11.31)
Total	37.43 (34.10)	19.67 (16.05)	32.86 (30.85)	16.86 (13.76)

Note. The values after an arcsine transformation are shown in parentheses.

Table 4.15

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Recall Rates of Explicit Text Information in Experiment 4

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	1802.26	1	1802.26	5.45	.026	.14
Error (P)	10917.35	33	330.83			
			Within-participants			
Text Type (T)	188.09	1	188.09	2.53	.121	.07
T × P	21.21	1	21.21	0.29	.597	.01
Error (T)	2452.32	33	74.31			
Total	15381.24	69				

4.1.4 Discussion

RQ4-1: Do Japanese EFL readers suppress the activation of predictive inferences when the subsequent context disconfirms the inferences?

The results of inference activation scores indicated that the strength of inference activation was not different between the predictive and disconfirming text conditions, regardless of readers' L2 proficiency. Therefore, the inferential information was not suppressed, and remained active immediately after the disconfirming context. Potts et al. (1988) found a similar result with L1 English readers; lexical and naming times for target words were not significantly different between predictive and disconfirming text conditions.

However, a study by Iseki (2006) with L1 Japanese readers yielded an opposite result to that of the present experiment; the meaningful judgment times for the target sentences were shorter for the predictive texts than the disconfirming texts, suggesting that inference activation was suppressed after the inferences were disconfirmed. One possible reason behind the different patterns of results among studies is variations in the tasks used to access inference activation. Importantly, Iseki pointed out the possibility that the task characteristics affected the obtained results: 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 response times for a target sentence (e.g., meaningfulness judgment times in Iseki, 2006) may be a better index of fluctuation in inference activation than those for a single target word (e.g., recognition times in the present experiment; lexical decision or naming times in Potts et al., 1988). Given this possibility, the suppression of predictive inferences during Japanese EFL learners' reading requires further examination using a task different from that used in this experiment prior to drawing a conclusion.

RQ4-2: Do Japanese EFL readers delete the disconfirmed predictive inferences from long-term text memory?

The results of the cued recall task found that the inferential information was recalled in the disconfirming text condition to the same degree as in the predictive text condition, and this was consistent between different proficiency groups. This suggests that the drawn inferences were not deleted but maintained in long-term text memory even when passages included the disconfirmation.

However, this result does not necessary mean that readers found disconfirmed inferences to be what actually happened in the narrative. A closer look at the participants' protocols in the disconfirming text condition indicated that some of the participants recalled target inferences as if they falsely believed that the inferred events actually happened in the described situations (e.g., *After finishing his work, he went to the lake and <u>drew a picture with pencils</u>*; translated from Japanese; inferential information is underlined), whereas others referred to their inferences simply to explain the character's actions (e.g., *She went to the jewelry shop to <u>sell the ring</u> her husband gave her, and she was hired at the shop*; translated from Japanese; inferential information is underlined). In the latter case, it was unclear whether the participants believed that the inferred event was what actually occurred. This is the limitation of recall analysis as a measure of the deletion of inferences. Therefore, further evidence is required in order to provide a more accurate picture of the deletion of disconfirmed inferences. Specifically, it is necessary to more directly assess whether readers believed that the inferred (but disconfirmed) events actually happened in the narratives.

RQ4-3: How does the disconfirmation of predictive inferences influence Japanese EFL readers' comprehension of explicit text information?

The results of the recall task also demonstrated that participants' recall performance for explicit text information was not significantly different regardless of text types. Inconsistent with Nahatame (2010), this suggests that the disconfirmation of the drawn predictive inferences had little impact on readers' understanding of explicit text information.

Nevertheless, the performance was numerically poorer in the disconfirming than the predictive text condition. In addition, a closer look at participants' recall protocols suggests that they constructed less coherent text representations in the disconfirming text condition. In some cases, participants included both inferential and disconfirming information in recall protocols as in the above example (e.g., *sell the ring at the shop* vs. *be hired at the shop*). These statements are somewhat inconsistent as the motivation for character's actions, but interestingly, readers nevertheless maintained both inferential and disconfirming information in comprehension. In sum, although the effect of the disconfirmation of predictive inferences on explicit comprehension was not statistically significant, descriptive statistics and qualitative observation of recall data raise the possibility that explicit text comprehension was reduced, and constructed text representations were less coherent for the disconfirming texts.

4.1.5 Conclusion of Experiment 4

The primary purpose of Experiment 4 was to examine the suppression and deletion of disconfirmed predictive inferences among Japanese EFL readers. The analysis of recognition times and recall protocols suggested that the activation of inferences remained immediately after the context disconfirming the inferences (the answer to RQ4-1), and the disconfirmed inferences were maintained in long-term text memory (the answer to RQ4-2). These results were consistent between higher and lower proficiency readers.

However, as discussed above, some methodological issues remained in Experiment 4. First, the results pertaining to the suppression of inferences might be affected by the characteristics of the task employed. Second, there is a potential problem with recall analysis; it was difficult to conclusively determine whether readers believed that the inferred event was what actually occurred in the described situation. Therefore, the results obtained in this experiment need to be reexamined using other tasks.

Additionally, although the cloze test used in this experiment showed high reliability, it has been argued that the test is unclear in terms of what aspects of language ability are measured (see Alderson, 1979, 1980). Thus, if the participants' L2 reading proficiency was more directly measured, the effect of L2 proficiency on suppression and deletion of predictive inferences might appear.

Given these issues, an additional experiment needs to be conducted in order to further investigate suppression and deletion of disconfirmed predictive inferences. Therefore, Experiment 5 was conducted to reexamine the suppression and deletion of predictive inferences among Japanese EFL readers.

The second purpose of Experiment 4 was to investigate whether disconfirmation of predictive inferences affected readers' explicit text comprehension. The obtained results were less clear: Although the effect was not statistically significant, it is possible that the disconfirmation of the drawn predictive inferences had some negative impacts on readers' explicit text comprehension (the answer to RQ4-3). Therefore, this issue also requires further investigation and the subsequent experiments used several tasks to investigate this issue. However, the obtained results were quite mixed. Thus, rather than discussing this issue within a single experiment, it is recommendable to discuss the issue based on several types of data obtained from multiple experiments. Consequently, the effect of disconfirmation of predictive inferences on explicit text comprehension is generally discussed in Chapter 5.

4.2 Experiment 5: Reexamining the Revision of Predictive Inferences in EFL Reading

4.2.1 Purpose and Hypotheses

Experiment 4 suggested that Japanese EFL learners have difficulty suppressing the incorrect predictive inferences during reading and deleting those inferences from long-term text memory. However, as described above, Experiment 4 had some methodological limitations. Thus, Experiment 5 aimed to overcome these limitations and reexamine the suppression and deletion of predictive inferences in EFL reading. The following two Hs were established based on the results of Experiment 4:

H3: Japanese EFL readers do not suppress the activation of predictive inferences when the subsequent context disconfirms the inferences.

H4: Japanese EFL readers do not delete the disconfirmed predictive inferences from long-term text memory.

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 methods or approaches; such experiments "help determine if a finding is an outcome of adapting a specific task or reflects a more general phenomenon" (p. 77).

In Experiment 5, instead of a single-word recognition task, participants performed an MJT immediately after reading each passage. In this task, participants are required to decide whether target sentences describing inferred events make sense or are semantically acceptable (Iseki, 2006). As mentioned in the discussion of Experiment 4, the MJT may be more sensitive to fluctuations in inference activation than the word recognition task because there is a greater degree of overlap between a target and an inferential concept. If the MJT is used and the results are different than those of Experiment 4, 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 Experiment 4, converging evidence from the different tasks strengthens the validity of the previous findings.

In addition, in Experiment 5, an SRT was used as an off-line measure instead of cued recall. As noted earlier, it was difficult to determine whether learners believed that the initially inferred (but later disconfirmed) event was what actually occurred based on the recall protocol analysis. To address this issue, Experiment 5 attempted to directly investigate the deletion of inferences by employing the SRT. In this task, after reading several experimental passages, participants are presented with target sentences describing inferred events in Japanese and then asked to judge whether the sentences are written in the passages (see also 2.3.2.2 for a review). If participants read the disconfirmed texts and believe that the initially inferred event did not actually happen in the described situation, the target sentence should be recognized as not written in the text.

Finally, instead of the cloze test, a multiple-choice L2 reading test was used in this experiment. Multiple-choice tests have been one of the most typical test types to measure L2 reading comprehension skill. Therefore, this test is assumed to more directly and precisely assess participants' L2 reading proficiency.

4.2.2 Methods

4.2.2.1 Participants

The participants of this experiment were 43 students (26 female and 17 male; aged 18–25 years, *M* = 20.49, *SD* = 1.79) from the same Japanese university. They were in different years (first o fourth) and had different majors (e.g., humanities, medical science etc.). All participants had studied English as a foreign language for more than six years as part of formal Japanese education, and they were assumed to have intermediate-level English proficiency.

The data from seven participants who did not follow the instructions or had an error rate that exceeded 30% on an experimental task (i.e., comprehension questions or the MJT) were omitted.[20] Therefore, the data analyses were based on 36 participants.

4.2.2.2 Materials

Passages, target sentences for the MJT, and comprehension questions

Thirty-two short narrative passages were adopted from prior research (Calvo & Castillo, 2001; Klin, Murray, et al., 1999; Motyka Joss, 2010; Virtue et al., 2006). As in Experiment 4, these passages were revised so that each of them consisted of four sentences comprised of simple words. These passages included 16 inference (i.e., predictive and disconfirming) texts and 16 neutral texts. Ten of the inference passages were previously used in Experiment 4, and six passages were newly constructed specifically for this experiment.[21] These inference passages suggest a specific motivational inference strongly constrained by the context (i.e., the MSC version). The neutral texts were similar to those in Experiment 4. Tables 4.16 and 4.17 show descriptive statistics for the experimental passages and disconfirming sentences used in Experiment 5, respectively. The number of words, lexical frequency of disconfirming sentences, and semantic relatedness to the fourth sentences were counterbalanced across four material sets, and there were no significant differences between sets (all Fs < 1.36). In addition to experimental passages, 16 filler passages were used.

[20] These included one participant who did not follow instructions, one participant who made more errors (above 30%) on the comprehension questions, four participants who made more errors (above 30%) on the MJT, and one participant who made more errors on both comprehension questions and the MJT.

[21] Six passages from Experiment 4 were not used because of difficulties in creating an appropriate MJT target sentence for each passage.

Table 4.16

Number of Words and the Readability of the Experimental Passages in Experiment 5

	M	*SD*	*Max*	*Min*	FKGL	FRE
Predictive text	52.88	9.95	71	36	5.29	79.26
Disconfirming text	67.69	10.51	91	51	5.39	80.45
Neutral text	41.87	8.18	62	30	4.18	77.55

Note. FKGL = mean Flesch-Kincaid Grade Level; FRE = mean Flesh Reading Ease.

Table 4.17

Number of Words, Lexical Frequency, and Semantic Relatedness for the Disconfirming Sentences in Experiment 5

	M	*SD*	*Max*	*Min*
Number of words	17.38	1.67	20	15
Lexical frequency	1.47	0.42	2.73	1.05
Semantic relatedness to fourth sentences	0.48	0.18	0.83	0.24

Note. Lexical frequency is the mean *JACET 8000* level of words included in the sentence. Semantic relatedness was calculated using Latent Semantic Analysis and ranged from 0 to 1.

Four material sets, each of which contained 12 experimental passages and 12 filler passages, were created to counterbalance the passages across versions. This ensured that each participant read an equal number of experimental passages of each type (i.e., predictive, disconfirming, and neutral), and each type of each passage was presented to an equal number of the participants.

Each experimental passage was paired with a target sentence for the MJT (MJT Target) and a simple comprehension question. Each MJT Target simply described the future event suggested by the predictive text (e.g., *He hits the ball*), including a main action verb directly representing a target inference concept (e.g., *hit*). All of the MJT Targets were composed of four or five words. Furthermore, these target sentences were written in the present tense and active voice to avoid the effects of participants' grammatical knowledge on the meaningfulness judgment. MJT Targets were also created for 12 filler passages, and all of them were semantically unacceptable sentences to balance the yes/no responses, though they included words related to the text content (e.g., *They drink the hamburger*).

The main verbs of the MJT Targets included three to five letters, had familiarity ratings greater than five on a 7-point scale (Yokokawa, 2006), and appeared in the most frequent 2,000 word level (Levels 1 and 2) in the *JACET List of 8000 Basic Words* (JACET Committee of Revising the JACET Basic Words, 2003). In addition, the number of letters and syllables, and lexical frequency of MJT Targets were counterbalanced across four material sets, and there were no significant differences (all Fs < 2.38). Table 4.18 shows descriptive statistics for the target sentences and main verbs included in these sentences. All of the MJT Targets are presented in Appendices 1 and 3 (see also Appendix 9 for their profiles).

As in previous experiments, comprehension questions were yes/no questions about an explicitly stated piece of text information. These questions were included to ensure the participants engaged in the reading.

Table 4.18

Length, Frequency, and Familiarity of MJT Targets

Target sentence	*M*	*SD*	*Max*	*Min*
Number of words	4.38	0.50	5	4
Number of syllables	5.31	1.14	8	4
Number of letters	17.69	2.94	23	13
Lexical frequency	1.36	0.57	2.60	1.00
Main verb				
Number of letters	4.25	0.77	5	3
Frequency (JACET 8000 Level)	1.06	0.25	2	1
Familiarity (on a 7-point scale from Yokokawa, 2006)	5.99	0.51	6.77	5.14

Note. Lexical frequency is the mean *JACET 8000* level of words included in the target sentence.

Target sentences for the SRT

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 (Muramoto, 2000). Table 4.19 shows an example of three target types for the inference text described in Table 4.1. All of the SRT Targets used in this experiment are shown in Appendix 7.

Table 4.19

Examples of the Three Types of Target Sentences for the SRT

Explicit	コーチは少年達の技能をテストすることにした [*The coach decided to test the boys' skills.*]
Inference	少年は球を打った [*The boy hit the ball.*]
Inconsistent	少年はグローブを磨いた [*The boy cleaned a baseball glove.*]

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.

SRT Targets were written also for four filler passages. The targets for each passage, however, included two explicit sentences and one inconsistent sentence to balance the number of yes and no responses of the SRT.

L2 reading proficiency test

A multiple-choice L2 reading test, adopted from Experiment 3, was used in this experiment. This test was created based on the reading subsection of retired copies of the EIKEN test, and included five passages with a total of 20 items.

4.2.2.3 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.

During the on-line phase of the experimental session, participants were randomly assigned to one of four material sets. Participants followed the same procedures as in

Experiment 4, except that immediately after reading each passage they performed the MJT instead of the recognition task. Participants were required to determine whether the presented 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. The passages were presented in random order.

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.

Following the on-line phase, the participants received one of other four material sets for the off-line phase. Each participant received the set that included 12 inference passages and four filler passages the participant had not read in the on-line phase, all presented on paper. The participants were asked to read these passages within 10 minutes. 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 texts they had just read, and (b) how confident they were of their judgment on a 4-point scale.

4.2.2.4 Scoring and Analysis

MJT response data as well as comprehension question data were eliminated if participants mistakenly skipped a sentence in the passage. In addition, response times greater than 2.5 standard deviations above the mean for each participant were excluded. This resulted in the exclusion of 2.78% of the MJT response data set. As in previous experiments, inference activation scores were calculated by subtracting the mean correct response times for the inference texts from those for the neutral texts. If readers activate target inferences during reading, correct responses to MJT Targets should be facilitated 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 (see Iseki, 2006; Potts et al., 1988; 2.3.3, p. 53).

The data from the SRT was analyzed based on Muramoto (2000). The recognition scores were calculated from participants' responses to SRT Targets and their confidence rating (see Table 2.4 in Chapter 2). The important focus is the recognition scores 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, possibly with high confidence, resulting in higher recognition scores. On the other hand, if readers initially activated the inferences but later delete them in their text memory, inference sentences are likely to be correctly recognized as not written in the text, possibly with high confidence, resulting in lower recognition scores.

4.2.3 Results

4.2.3.1 L2 Reading Proficiency Test

The reliability of the L2 reading proficiency test was sufficient (Cronbach's $\alpha = .78$) 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 ($n = 18$) and a lower proficiency group ($n = 18$). There was a significant difference in test scores between these two groups, $t(24.44) = 8.01$, $p < .001$, $d = 2.67$. Table 4.20 shows the mean proficiency test scores for each proficiency group.

Table 4.20

Mean Scores on the L2 Reading Proficiency Test in Experiment 5

	M	*SD*	*Max*	*Min*
Higher (*n* = 18)	13.50	1.25	17	12
Lower (*n* = 18)	8.06	2.60	11	3

Note. Maximum possible score is 17.

4.2.3.2 Comprehension Questions

The overall average percent correct for the comprehension questions in the on-line phase was quite high (above 90%), supporting the claim that participants engaged in the text reading. Table 4.21 shows the mean correct response rates (%) for each text type and proficiency group.

Table 4.21

Mean Accuracy (%) on the Comprehension Questions for Experimental Passages in Experiment 5

	Predictive		Disconfirming		Neutral	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	95.83	9.59	93.98	11.72	92.59	12.42
Lower	92.59	12.42	94.44	10.69	87.96	16.47
Total	94.21	11.06	94.21	11.06	90.28	14.57

A 3 (Text Type: predictive, disconfirming, neutral) × 2 (Proficiency: higher, lower) mixed ANOVA on the means for the experimental passages did not indicate significant main effects of Text Type and Proficiency; the interaction was not statistically significant (see Table 4.22).

Table 4.22

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Mean Comprehension Question Accuracy in Experiment 5

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.02	1	0.02	1.22	.277	.03
Error (P)	0.46	34	0.01			
			Within-participants			
Text Type (T)	0.04	2	0.02	1.14	.327	.03
T × P	0.01	2	0.01	0.38	.684	.01
Error (T)	1.11	68	0.02			
Total	1.64	107				

4.2.3.3 Meaningfulness Judgment Times and Accuracy

For the MJT, the mean percent correct was 92.68% (*SD* = 4.80) for all passages including fillers, 90.26% (*SD* = 6.75) for the experimental passages only. Table 4.23 shows the response times for the correct responses in each text condition. Figure 4.2 represents inference activation scores calculated from the correct response times.

As in Experiment 4, it is necessary to confirm that target inferences were activated when participants read the predictive text before examining the suppression of inferences. Two-tailed paired *t* tests were conducted on mean inference activation scores for the predictive text condition to determine whether the activation scores were greater than zero. The activation scores for all readers were significantly greater than zero, $t(35) = 3.67$, $p = .001$, $d = 0.60$. Similarly, those scores for both higher and lower proficiency readers were significantly greater than zero, $t(17) = 3.66$, $p = .002$, $d = 0.84$; $t(17) = 2.34$, $p = .032$, $d =$

0.54. Thus, it was confirmed that participants activated the expected inferences when they read the predictive context.

Table 4.23

Mean Correct Response Times (ms) for the MJT

	Predictive		Disconfirming		Neutral	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	1,793	390	1,915	530	2,121	530
Lower	2,223	476	2,325	807	2,688	906
Total	2,008	481	2,120	705	2,405	786

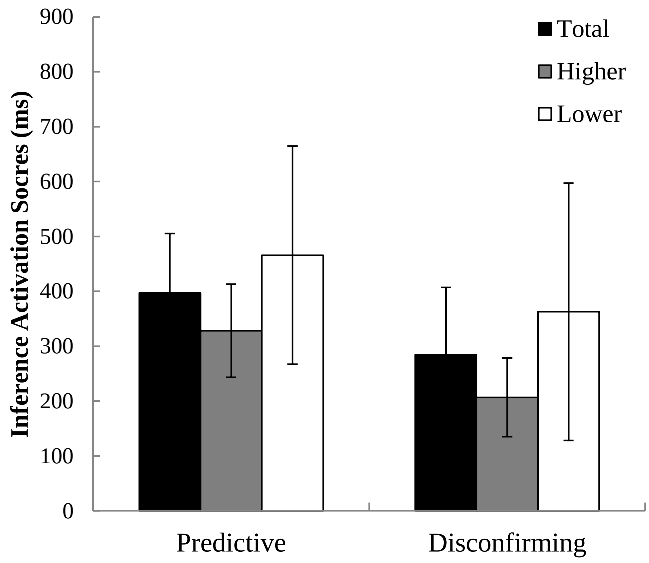

Figure 4.2. Mean inference activation scores (ms) as a function of text type and L2 proficiency in Experiment 5. Error bars represent the standard errors of the means. Higher × Predictive: *M* = 328.13, *SD* = 380.09; Lower × Predictive: *M* = 465.79, *SD* = 844.08; Higher × Disconfirming: *M* = 206.59, *SD* = 320.69; Lower × Disconfirming: *M* = 362.68, *SD* = 995.22.

To examine the suppression of the activated predictive inferences, a 2 (Text Type: predictive, disconfirming) × 2 (Proficiency: higher, lower) ANOVA was conducted on the mean inference activation scores, with Text Type as a within-participants variable and Proficiency as a between-participants variable. The results showed that the main effects of Text and Proficiency, and a Text Type × Proficiency interaction were not statistically significant (see Table 4.24). Thus, there were no significant differences in activation scores between the predictive and disconfirming conditions regardless of participants' L2 reading proficiency.

Table 4.24

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Inference Activation Scores in Experiment 5

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	388271.41	1	388271.41	0.46	.503	.01
Error (P)	28802750.47	34	847139.72			
			Within-participants			
Text Type (T)	227102.05	1	227102.05	1.77	.193	.05
T × P	1527.57	1	1527.57	0.01	.914	.00
Error (T)	4351694.19	34	127991.01			
Total	33771345.69	71				

As described above, the overall mean accuracy for the MJT was high, and the error rates were low in each condition (see Table 4.25). Any significant effects were not found in a 2 (Text Type: predictive, disconfirming) × 2 (Proficiency: higher, lower) mixed ANOVA on response accuracy facilitation (i.e., differences between inference and neutral texts) (higher × predictive: M = 1.30, SD = 20.59; lower × predictive: M = 1.83, SD = 16.05; higher × disconfirming: M = 5.11, SD = 19.17; lower × disconfirming: M = −0.94, SD = 18.90) (see Table 4.26). Therefore, these data indicate that there was no speed-accuracy trade-off in the MJT.

Table 4.25

Mean Response Accuracy (%) for the MJT

	Predictive		Disconfirming		Neutral	
Proficiency	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	90.28	15.19	93.98	11.72	88.89	12.78
Lower	91.67	12.13	88.89	12.78	89.81	13.27
Total	90.97	13.57	91.44	12.36	89.35	12.85

Table 4.26

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Facilitation of Meaningfulness Judgment Accuracy

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.01	1	0.01	0.28	.597	.01
Error (P)	1.69	34	0.05			
			Within-participants			
Text Type (T)	0.00	1	0.00	0.02	.887	.00
T × P	0.02	1	0.02	0.98	.329	.03
Error (T)	0.66	34	0.02			
Total	2.39	71				

4.2.3.4 Sentence Recognition

The mean recognition scores for each sentence type and text condition are reported in Table 4.27 and Figure 4.3.

Table 4.27

Mean Recognition Scores for the SRT in Experiment 5

	Explicit				Inference				Inconsistent			
	Pre		Dis		Pre		Dis		Pre		Dis	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	4.94	0.82	4.99	1.02	4.19	1.18	2.39	1.07	0.92	0.55	1.05	0.76
Lower	5.06	0.75	4.79	1.20	3.97	1.40	2.85	1.31	1.28	1.03	1.11	0.85
Total	5.00	0.78	4.89	1.10	4.08	1.28	2.62	1.20	1.10	0.83	1.08	0.80

Note. The scores ranged from 0 to 6. Pre = predictive; Dis = disconfirming.

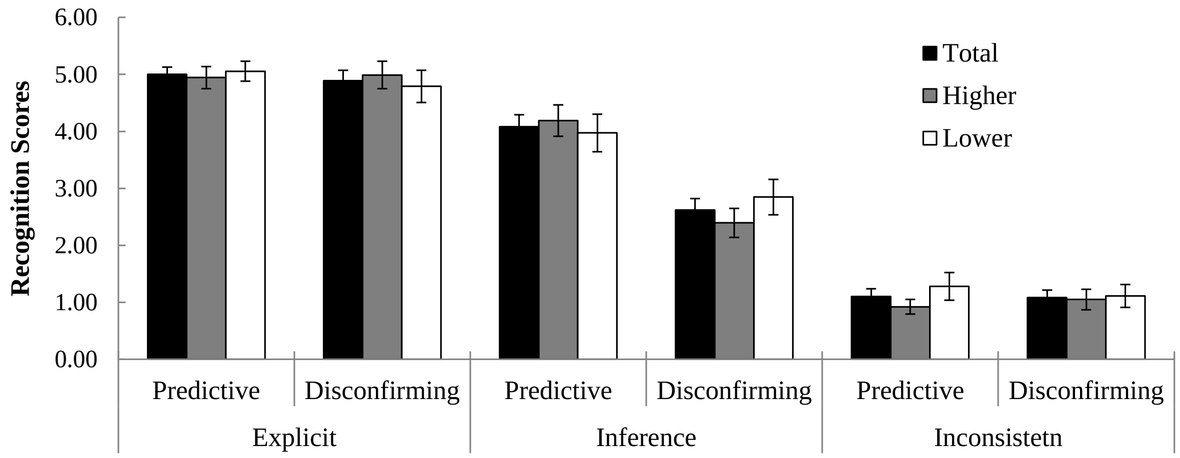

Figure 4.3. Mean recognition scores for the SRT in Experiment 5. Error bars represent the standard errors of the means.

To examine the deletion of inferences, a 3 (Sentence Type: explicit, inference, inconsistent) × 2 (Text Type: predictive, disconfirming) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on the recognition scores, with Sentence and Text as within-participants variables and Proficiency as a between-participants variable. The ANOVA revealed significant main effects of Sentence and Text, and a significant Sentence × Text interaction. None of the other effects were significant or marginally significant (see Table 4.28).

Table 4.28

Summary Table for Three-Way ANOVA of the Effects of Proficiency, Sentence Type and Text Type on Recognition Scores in Experiment 5

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.48	1	0.48	0.48	.495	.01
Error (P)	34.15	34	1.00			
			Within-participants			
Sentence Type (S)	540.39	2	270.19	168.23	.000	.83
S × P	0.59	2	0.29	0.18	.833	.01
Error (S)	109.21	68	1.61			
Text Type (T)	15.13	1	15.13	29.56	.000	.47
T × P	0.01	1	0.01	0.01	.912	.00
Error (T)	17.40	34	0.51			
S × T	23.50	2	11.75	14.96	.000	.31
S × T × P	2.85	2	1.43	1.82	.171	.05
Error (S × T)	53.41	68	0.79			
Total	762.50	180				

Follow-up tests of the significant interaction revealed that there were significant differences between all sentence types for both predictive and disconfirming conditions (see Table 4.29). More importantly, the simple main effect of Text was significant for inference sentences, but not for explicit and inconsistent sentences. The scores for inference sentences were significantly higher in the predictive text condition than the disconfirming text 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 text condition suggest that participants recognized these sentences as not written with a low to relatively low confidence level (see Table 2.4 in Chapter 2).

Table 4.29

Summary Table for Follow-Up Tests on Recognition Scores in Experiment 5

			p	*d*
	[Explicit]	[Inference]	.001	0.87
Predictive	[Explicit]	[Inconsistent]	.000	4.84
	[Inference]	[Inconsistent]	.000	2.76
	[Explicit]	[Inference]	.000	1.97
Disconfirming	[Explicit]	[Inconsistent]	.000	3.97
	[Inference]	[Inconsistent]	.000	1.51
Explicit	[Predictive]	[Disconfirming]	.610	0.12
Inference	[Predictive]	[Disconfirming]	.000	1.18
Inconsistent	[Predictive]	[Disconfirming]	.905	0.02

4.2.4. Discussion

H3: Japanese EFL readers do not suppress the activation of predictive inferences 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 after processing the disconfirming context, as indicated by the lack of a significant difference in the activation scores of the text conditions. This suggests that inference activation was not suppressed even when the subsequent context disconfirmed the inferences. These results are highly consistent with those of Experiment 4, which used a word recognition 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 Japanese EFL readers have difficulty suppressing the activation of predictive inferences immediately after reading a disconfirming sentence.

It is important to consider why these readers had trouble suppressing predictive inferences. Iseki (2006) used the MJT as in the present experiment, and demonstrated that predictive inference activation was suppressed by L1 Japanese readers. 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. However, it should be noted that the experimental passages used in Experiments 4 and 5 were relatively short and simple so that the participants could easily comprehend them. In addition, regardless of participants' L2 proficiency level as assessed by the cloze test and multiple-choice reading test, inference activation was not suppressed. Taken together, this finding is attributable to a more direct influence of other factors rather than the cognitive demands of L2 reading.

One such factor is the characteristics of text information disconfirming the inferences. Guéraud et al. (2005) indicated that although readers' comprehension difficulties occur when previously-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. In the present experiments, the information disconfirming the target inference was described in just one sentence (including three to four IUs), whereas the context preceding inference activation consisted of four sentences (including five to 10 IUs). Although the disconfirming context suggested that the activated inferences were incorrect, it seemed as though the amount of context was insufficient to fully suppress the inferences. Furthermore, the disconfirming context neither explicitly disconfirmed the drawn inference nor strongly encouraged readers to generate an alternative inference instead of the disconfirmed one. Previous studies found that the difficulty of integrating inconsistent statements is reduced when the context explicitly negates the information or provides a clear reason for the revision of previously stated information (Kendeou et al., 2013; O'Brien et al., 1998; Rapp & Kendeou, 2009). Therefore, the quality of the disconfirming context as well as the quantity could be a factor affecting the difficulty of suppression.

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 <u>to hit the ball</u>*). 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. Moreover, in some passages, 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.

The present results also indicated that, consistent with Experiment 4, the activation of inferences was not suppressed by either higher or lower proficiency readers. This seems contrary to the SBF assumption that more skilled readers have efficient suppression mechanisms (Gernsbacher, 1990; Gernsbacher et al., 2004). One of the reasons for the lack of L2 proficiency effects could be that the impact of the aforementioned text and target inference characteristics was so strong that the effects of L2 proficiency became less clear. Another possibility is that the mechanism supposed by the SBF, which concerns the suppression of explicit text information during L1 reading, is not necessarily applicable to the suppression of inferential information during L2 reading. Consequently, more data are required to make conclusions about the effects of L2 proficiency on the suppression of predictive inferences during EFL reading.

The results of the MJT suggested that immediately suppressing the activation of inferences during reading was difficult for the participants in this experiment. However, they need to at least understand that the inferred event did not actually happen in the situation described by disconfirming passages. If they maintain disconfirmed inferences in text memory and falsely believe that inferred event actually happened, they should fail to accurately comprehend the text. The next section discusses this issue.

H4: Japanese EFL readers do not delete the disconfirmed predictive inferences from long-term text memory.

The implication of the SRT results somewhat differ from that of the recall task results of Experiment 4. The statistical analysis of the recognition scores revealed significant differences between the predictive and disconfirming text conditions only for inference sentences, with higher scores in the predictive than in the disconfirming text condition. This suggests that readers delete the target predictive inferences form their memory after they read the disconfirming context.

However, it should be noted that the recognition scores in the disconfirming text condition were significantly higher for inference than inconsistent sentences. Furthermore, the mean recognition scores suggested that the participants were less confident in their judgment when they rejected the inference sentences in the disconfirming text condition. Taken together, the most plausible interpretation of the SRT results is that the participants partially, but not completely, deleted the disconfirmed inferences. In other words, although Japanese EFL readers tend to delete the predictive information after the context disconfirms it, it can be difficult to eliminate the memory trace of the strongly predicted information. It is apparent that several factors mentioned in the discussion of H3 (i.e., the quantity and quality of disconfirming context and characteristics of target inferences) are also likely to cause difficulty in deleting disconfirmed inferences from memory.

Furthermore, when the pattern of results for the MJT is compared to that of the SRT, there appears to be an inconsistency. The participants found it difficult to suppress their inferences during reading, but they partially deleted predictive inferences after reading. As in Experiment 2, there are some cases where the process of reading comprehension is not reflected in the product of reading comprehension (Horiba, 2013). Nevertheless, 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 reread the earlier context. In other words, it is possible that the sentence-by-sentence presentation of the text prevented participants from looking back to the previous context during reading, which increased the cognitive load on reading and prohibited the reanalyzing of activated inferences based on the earlier context. This might result in difficulty suppressing disconfirmed inferences.

4.2.5 Conclusion of Experiment 5

Experiment 5 examined the suppression and deletion of incorrect predictive inferences with an awareness of methodological issues in evaluating the suppression and deletion of predictive inferences. The results of this experiment confirmed and extended earlier findings. First, consistent with Experiment 4, the results of the MJT suggested that Japanese EFL readers have difficulty suppressing the activation of predictive inferences immediately after reading the context disconfirming the inferences (support for H3). In contrast, the SRT results revealed that they achieved a partial deletion of inferences in text memory based on the disconfirming context, although it was still difficult to complete the deletion of the inferences (partial support for H4).

However, some methodological limitations remain in Experiment 5. First, participants read experimental passages sentence-by-sentence on the computer screen before the response task. As described above, this prevented participants from looking back to prior context during reading, which might increase the cognitive load on reading compared to natural reading.

Second, because the response task was conducted after reading the sentence disconfirming the inferences, it provided little information about how participants processed the sentence. Consequently, it is difficult to determine what kinds of comprehension processes are related to the difficulty of suppressing and deleting disconfirmed inferences. For instance, given that L2 readers are more likely to dismiss inconsistencies included in the passages compared to L1 readers (Morishima, 2013), it is possible that the Experiment 5 participants dismissed inconsistencies between drawn inferences and the sentence being processed, or that there was insufficient awareness of the inconsistences. Alternatively, participants might have more difficulty integrating the disconfirming sentence with the comprehension of prior context, even though they were aware of inconsistencies, similar to poor L1 readers (Long & Chong, 2001). However, examining these possibilities requires the use of measures that can tap into real-time processing of the sentence disconfirming the inferences.

To overcome these limitations of Experiment 5, it will be helpful to adopt the eye tracking methodology (see 2.3.2.2 for a review). Therefore, Experiment 6 investigated learners' text comprehension processes when they encounter the context disconfirming the inferences by collecting data about their eye movements during reading.

4.3 Experiment 6: Revision of Predictive Inferences and Text Comprehension Processes in EFL Reading

4.3.1 Purpose, Hypothesis, and Research Question

As described in the previous section, methodological limitations of Experiment 5 can be overcome by the use of eye-tracking method. Specifically, it can examine learners' comprehension in a more natural reading setting in ways such as presenting an entire passage and not giving participants extra requirements (e.g., frequently pressing a button). Furthermore, it allows us to investigate the real-time processing of the sentence that disconfirms predictive inferences. It can also identify what comprehension processes are more related to difficulty revising inferences by discriminating early and late processes in text comprehension. For these reasons, eye tracking was adopted in Experiment 6.

The purpose of the Experiment 6 was to confirm and extend the findings of Experiment 5. First, to refine the previous finding that learners have difficulty eliminating disconfirmed inferences from text memory, their text memory was examined using the same recognition task used in Experiment 5. Second, eye movement data during reading were collected to examine comprehension processes when readers encounter the sentence that disconfirms predictive inferences. Consequently, the following H and research RQ were addressed:

H5: Japanese EFL learners have some difficulty eliminating disconfirmed predictive inferences from text memory.

RQ4-4: How do the inconsistencies between drawn predictive inferences and the following text affect Japanese EFL learners' text comprehension processes?

The present experiment examined participants' eye movements using a similar methodology to previous studies on inconsistency processing (e.g., Hyönä et al., 2003; Rinck et al., 2003). Specifically, it analyzed eye movements on the sentence disconfirming predictive inferences drawn from the prior context (see 4.3.2.4 for further explanations). As in previous experiments, the present experiment included learners' L2 reading proficiency as a reader-related variable.

4.3.2 Methods

4.3.2.1 Participants

The participants were 26 Japanese university students (16 male and 10 female; aged 18–22 years, M = 19.96, SD = 1.46). They were in different years (first to fourth) and had different majors (e.g., international studies, social science, medical science etc.). All participants were native speakers of Japanese with normal or corrected-to-normal visual acuity. In addition, participants had studied English as a foreign language for more than six years as part of Japanese formal education, and they were assumed to have intermediate-level English proficiency.

The data from eight participants was excluded from analyses because of inaccurate eye-movement recordings. Therefore, the following analyses were based on 18 participants.

4.3.2.2 Materials

Passages and comprehension questions

Twelve disconfirming texts were adopted from Experiment 5. Each passage consisted of five sentences. The fourth sentence strongly induces a single, specific predictive inference (i.e., target inference), but this inference is disconfirmed by the fifth sentence. Table 4.30 shows descriptive statistics for the disconfirming sentences used in Experiment 6,

Table 4.30

Number of Words and Lexical Frequency for the Disconfirming Context in Experiment 6

	M	*SD*	*Max*	*Min*
Number of words	17.50	1.68	20	15
Lexical frequency	1.48	0.46	2.73	1.05

Note. Lexical frequency is the mean *JACET 8000* level of words included in the sentence. Semantic relatedness was calculated using Latent Semantic Analysis and ranged from 0 to 1.

The present experiment constructed new control passages corresponding to each disconfirming passage. The control texts were created by rewriting the fourth sentence of each disconfirming text so that it did not induce the target inference and naturally precedes the content of the fifth sentence (see Table 4.31 for an example; see also Appendix 5 for all the control passages used in Experiment 6). Therefore, in the control texts, the fifth sentences do not disconfirm the participants' predictions, although the sentences were the same as those in the disconfirming texts.

Table 4.31

Sample of Experimental Passages in Experiment 6

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. <u>As the pitcher released the ball, the boy raised his bat and the ball went directly towards him.</u> (Disconfirming) / <u>As the pitcher released the ball, the boy raised his bat without knowing it was a forkball</u>. (Control) Suddenly, the ball dropped in front of the bat and fell in the catcher's mitt.

Table 4.32 shows descriptive statistics for the experimental passages used in Experiment 6. The number of words and readability were not significantly different between disconfirming and control passages (all *t*s < .1.84). Similarly, the number of words and lexical frequency were counterbalanced across two material sets, and there were no significant differences between sets (all *t*s < 1.44).

Table 4.32

Number of Words and the Readability of the Experimental Passages in Experiment 6

	M	*SD*	*Max*	*Min*	FKGL	FRE
Disconfirming text	70.00	5.66	91	51	5.66	78.88
Control text	70.25	5.99	96	50	5.97	77.82

Note. FKGL = mean Flesch-Kincaid Grade Level; FRE = mean Flesh Reading Ease.

The plausibility of the fourth sentences of control texts was confirmed and improved in a small pilot study with 12 Japanese graduate students majoring in English education. Furthermore, the fourth sentences of the control texts were not significantly different than those of the disconfirming texts in terms of the sentence length (i.e., the number of words), lexical frequency (JACET 8000 level), and semantic relatedness to the fifth sentences, all *t*s < 1.77. Table 4.33 shows the descriptive statistics for fourth sentences of disconfirming and control texts

Table 4.33

Number of Words, Lexical Frequency, and Semantic Relatedness for Fourth Sentences of Experimental Passages

	Number of words		Lexical frequency		Semantic relatedness	
	M	*SD*	*M*	*SD*	*M*	*SD*
Disconfirming	15.67	3.55	1.48	0.32	0.49	0.20
Control	16.17	3.46	1.45	0.40	0.45	0.21

Note. Lexical frequency is the mean JACET 8000 level of words included in the sentence. Semantic relatedness was calculated using Latent Semantic Analysis.

In addition to experimental passages, six filler passages were used to prevent participants from becoming aware of the purpose of the experiment. The filler passages were similar in length to the experimental passages, but neither elicited nor disconfirmed any specific inferences.

As in previous experiments, all experimental and filler texts were paired with yes/no comprehension questions. These questions concerned a piece of information explicitly stated in the text before the fifth sentences.

Two material sets were constructed, resulting from the counterbalance of passage types. Each list included six disconfirming, six control, and six filler passages. This ensured that each participant read an equal number of experimental passages of each type, and that each type of experimental passage was presented to one half of the participants.

Target sentences for the SRT

Target sentences for the post-reading recognition task were also adopted from Experiment 5. Each text was paired with explicit, inference, and inconsistent sentences. Each target sentence was presented with a 4-point scale of recognition judgment confidence (1 = *low*, 2 = *relatively low*, 3 = *relatively high*, and 4 = *high*), similar to Experiment 5.

L2 reading proficiency test

The same multiple-choice test as in Experiment 5 was used.

4.3.2.3 Procedure

An L2 reading proficiency test was conducted 30 minutes prior to the main experimental session. In the main session, participants sat approximately 40cm from the computer screen displaying the texts. At this distance, two letters spaces equaled 1º of the visual angle. The participants' head position was fixed by means of a chinrest. Using the SuperLab 4.5 program, each text was displayed in its entirety on the computer screen. To make analysis of eye movements more accurate, each sentence started on a new line, and line spacing was inserted between sentences (see Figure 4.4). Eye movements during reading were recorded with the eye-tracker (EMR 9) manufactured by nac Image Technology Inc. (Tokyo, Japan).

After the eye-tracker was adjusted for optimal tracking, participants' eye movements were calibrated with a standard 9-point grid. In the subsequent experimental trials, participants were given one of the two material sets. Each trial began with "Ready" displayed on screen. Participants were asked to push the button on the Response Pad RB-730 (Cedrus, U.S.) to indicate they were ready to read a passage. Pushing the button caused the passage to appear on the screen. Participants then silently read the passage in a self-paced manner. They were instructed to press the button again when they finished the reading. Subsequently, the current passage was replaced with the comprehension question. Participants responded to the question using a pair of yes/no buttons, and received feedback about accuracy. This trial was repeated for each of the 18 passages presented in random order. Before starting the experiment, participants completed practice trials for three passages.

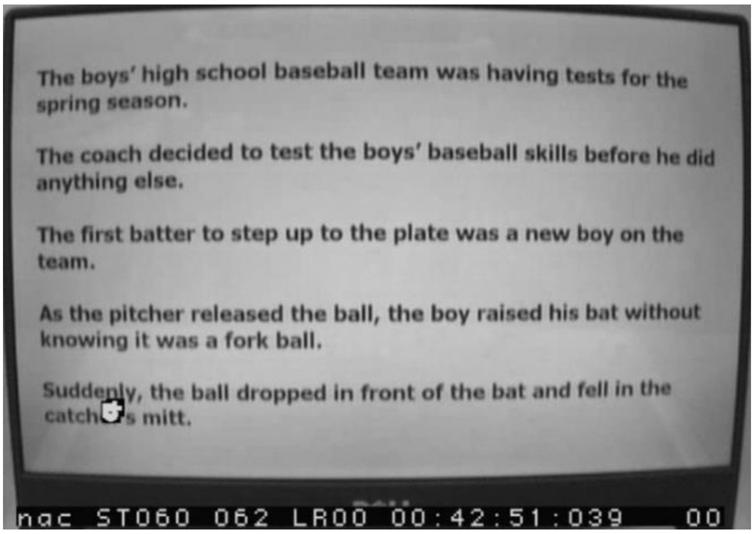

Figure 4.4. Sample of the text display and eye movement recordings.

After reading all passages, participants received the booklet that included target sentences for the recognition task. The booklet presented all three types of target sentences for each text they had just read. Participants were instructed to: (a) judge whether each sentence was described in the text they had read, and (b) rate how confident they were of their judgments on a 4-point scale.

4.3.2.4 Scoring and Analysis

The data from the SRT was analyzed in the same way as in Experiment 5. The recognition scores, ranging from 0 to 6, were calculated from participants' responses to the target sentences and their confidence ratings (see Table 2.4).

Regarding the eye movement data, those from the selected eye (usually the right eye) were analyzed. All trials where recording problems occurred were omitted from analysis

(1.39% of all trials). This study analyzed fixation times on the fifth sentences of experimental passages (these sentences were common between passage types and defined as target sentences). Prior to analysis, short fixation durations of less than 100 ms were excluded, along with long fixation durations beyond three standard deviations from each participant's mean in a condition (1.63% of the data set). In the analysis, individual eye fixations within a target sentence were ignored, but their durations were summed to yield the fixation durations for the sentence (Hyönä et al., 2003).[22]

As in previous studies (e.g., Hyönä et al., 2003; Rinck et al., 2003), this study analyzed first-pass and second-pass reading times on target sentences. First-pass reading time is defined as the sum of the duration of all fixations on the target sentence until finishing reading or before looking back to prior sentences. This eye fixation measure reflects the initial processing of the sentence, and is indicative of the readers who noticed inconsistencies between drawn inferences and the sentence being processed. If readers consciously noticed the inconsistencies, such awareness should make initial processing of the target sentence more difficult, resulting in an increase of first-pass reading times (van der Schoot et al., 2012). On the other hand, if readers dismissed or did not fully notice the inconsistencies, little effect should appear on this measure.

Second-pass reading time was defined as the time of additional fixations on the target sentence that occurred after looking back to one or more preceding sentences. This eye fixation measure reflects later comprehension processes, such as the integration of the target sentence with comprehension of preceding context (i.e., developing representations). Thus, if readers have more integration difficulty of the target sentences, some effects are assumed to appear on this measure.

[22] Because the length of each target sentence varied across passages, this experiment analyzed raw fixation durations as well as those divided by the number of syllables (Rinck et al., 2003) and characters (van der Schoot et al., 2012) in the sentence. Since similar results were obtained by these analyses, this experiment reports only the results of raw fixation durations.

In addition to these eye fixation measures, sentence wrap-up time was also analyzed as the total fixation times on the final two words of the target sentence (van der Schoot et al., 2012). This was because second-pass reading on the target sentences did not frequently occur in this study (see 4.3.3.3.2). The processes during reading the end of the sentence (i.e., sentence wrap-up) include not only integrating words with the sentence being processed, but also integrating the sentence with the preceding context (Just & Carpenter, 1980). Thus, this eye fixation measure also indicates the discourse integration difficulty of target sentences.

4.3.3 Results

4.3.3.1 L2 Reading Proficiency Test

The reliability of the L2 reading proficiency test was sufficient (Cronbach's $\alpha = .81$) after excluding three low discriminability items. Table 4.34 shows the mean scores of the proficiency test for each proficiency group. Before the main analysis, participants were classified into either a higher ($n = 8$) or lower reading proficiency group ($n = 10$) based on a median split of the test scores. The mean scores were significantly higher for the higher proficiency group than the lower proficiency group, $t(16) = 7.76$, $p < .001$, $d = 3.68$.

Table 4.34

Mean Scores on the L2 Reading Proficiency Test in Experiment 6

	M	*SD*	*Max*	*Min*
Higher (*n* = 8)	15.00	1.67	17	12
Lower (*n* = 10)	8.60	1.84	11	6

Note. Maximum possible score is 17.

4.3.3.2 Sentence Recognition

Table 4.35 and Figure 4.5 show the mean recognition scores for the SRT. A 3 (Sentence Type: explicit, inference, inconsistent) × 2 (Text Type: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on the mean scores, with Sentence Type and Text Type as within-participants variables and Proficiency as a between-participants variable. The results indicated significant main effects of Sentence Type and Text Type. In addition, there was a significant interaction between Sentence Type and Text Type. Any other potential main or interaction effects were not significant or marginally significant (see Table 4.36).

The subsequent analysis of the Sentence Type × Text Type interaction revealed a significant difference between disconfirming and control texts for inference sentences (see Table 4.37), with higher scores for disconfirming than control texts. There was no significant difference between texts for explicit and inconsistent sentences.

Table 4.35

Mean Recognition Scores for the SRT in Experiment 6

	Explicit				Inference				Inconsistent			
	Dis		Con		Dis		Con		Dis		Con	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	5.13	0.59	5.04	0.82	2.21	1.06	1.33	0.95	0.92	0.67	0.90	0.58
Lower	5.17	0.49	4.97	0.71	2.53	1.07	1.42	0.63	0.73	0.38	0.90	0.61
Total	5.15	0.52	5.00	0.74	2.39	1.05	1.38	0.76	0.82	0.52	0.90	0.58

Note. The scores ranged from 0 to 6. Dis = disconfirming passages; Con = control passages.

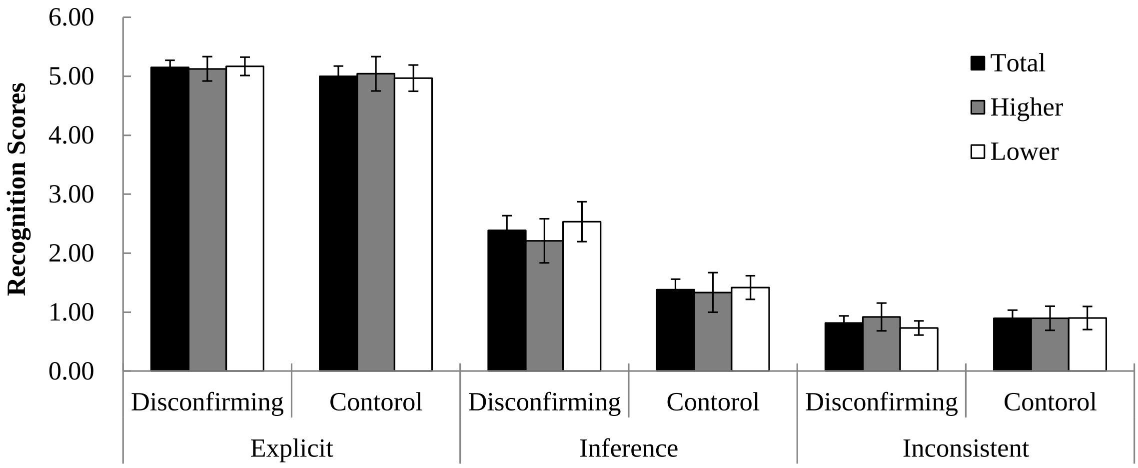

Figure 4.5. Mean recognition scores for the SRT in Experiment 6. Error bars represent the standard errors of the means.

Table 4.36

Summary Table for Three-Way ANOVA of the Effects of Proficiency, Sentence Type and Text Type on Recognition Scores in Experiment 6

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.03	1	0.03	0.04	.852	.00
Error (P)	12.74	16	0.80			
			Within-participants			
Sentence Type (S)	344.06	2	172.03	228.20	.000	.93
S × P	0.42	2	0.21	0.28	.760	.02
Error (S)	24.12	32	0.75			
Text Type (T)	3.36	1	3.36	11.97	.003	.43
T × P	0.02	1	0.02	0.08	.788	.00
Error (T)	4.50	16	0.28			
S × T	5.67	2	2.83	8.48	.001	.35
S × T × P	0.22	2	0.11	0.33	.724	.02
Error (S × T)	10.70	32	0.33			
Total	405.83	107				

Table 4.37

Summary Table for Follow-Up Tests on Recognition Scores in Experiment 6

			p	*d*
	[Explicit]	[Inference]	.000	3.33
Disconfirming	[Explicit]	[Inconsistent]	.000	8.36
	[Inference]	[Inconsistent]	.000	1.90
	[Explicit]	[Inference]	.000	4.82
Control	[Explicit]	[Inconsistent]	.000	6.18
	[Inference]	[Inconsistent]	.029	0.71
Explicit	[Predictive]	[Disconfirming]	.492	0.23
Inference	[Predictive]	[Disconfirming]	.000	1.10
Inconsistent	[Predictive]	[Disconfirming]	.527	0.15

4.3.3.3 Eye Movement Measures

4.3.3.3.1 First-Pass Reading Time

Table 4.38 shows mean first-pass reading time of target sentences for each text type and proficiency group. A 2 (Text Type: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA indicated significant main effects of Text Type and Proficiency (see Table 4.39). Specifically, mean reading time was significantly longer for disconfirming than control texts. Furthermore, mean reading time was significantly shorter for higher than lower proficiency learners. There was no significant interaction effect between Text Type and Proficiency.

Table 4.38

First-Pass Reading Times (ms) of Target Sentences

	Disconfirming		Control	
Proficiency	*M*	*SD*	*M*	*SD*
Higher	6,455	2,339	5,421	1,498
Lower	8,370	1,359	6,996	1,513
Total	7,519	2,047	6,296	1,669

Table 4.39

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on First-Pass Reading Times

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	27070669.85	1	27070669.85	5.77	.029	.26
Error (P)	75114913.17	16	4694682.07			
			Within-participants			
Text Type (T)	12892922.70	1	12892922.70	12.78	.003	.44
T × P	258005.68	1	258005.68	0.26	.620	.02
Error (T)	16140601.53	16	1008787.60			
Total	131477112.94	35				

4.3.3.3.2 Second-Pass Reading Time

Mean second-pass reading time of target sentences are shown in Table 4.40. A 2 (Text Type: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA found neither significant main effects of Text Type nor Proficiency (see Table 4.41). There was also no significant interaction effect between Text Type and Proficiency.

Table 4.40

Second-Pass Reading Times (ms) of Target Sentences

	Disconfirming		Control	
Proficiency	*M*	*SD*	*M*	*SD*
Higher	291	428	439	748
Lower	694	945	616	1,221
Total	515	769	537	1,014

Table 4.41

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Second-Pass Reading Times

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	651739.78	1	651739.78	0.67	.427	.04
Error (P)	15670826.61	16	979426.66			
			Within-participants			
Text Type (T)	44.71	1	44.71	0.00	.992	.00
T × P	154669.46	1	154669.46	0.33	.574	.02
Error (T)	7505505.08	16	469094.07			
Total	23982785.64	35				

However, it should be noted that the mean second-pass reading time was quite small for each condition, which corresponded to only a few fixations. This suggests that participants did not frequently reread the target sentences after regressions towards the prior sentences. In addition, the large standard deviations suggest that second-pass reading times greatly varied between individual participants. Indeed, the overall probability of second-pass reading on the target sentence was less than 20%, and seven of 18 participants made no second-pass reading on target sentences throughout all trials.[23] Therefore, it is difficult to interpret the results of second-pass reading time made by such a limited number of fixations and participants. Consequently, this study also analyzed the sentence wrap-up times as another measure reflecting discourse integration difficulty of target sentences in the following section.

4.3.3.3.3 Sentence Wrap-Up Time

Table 4.42 shows mean wrap-up times of target sentences. A 2 (Text Type: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA indicated the significant main effects of Text and Proficiency (see Table 4.43). More importantly, there was a significant interaction between Text Type and Proficiency. The subsequent analysis indicated that sentence wrap-up times among the higher proficiency group were not significantly different between disconfirming and control texts, $t(7) = 0.23$, $p = .821$, $d = 0.05$, whereas the lower proficiency group showed significantly longer times for disconfirming than control texts, $t(9) = 4.80$, $p = .001$, $d = 1.17$.

[23] Participants also did not frequently regress towards the context preceding the target sentence. The overall mean probability of making regressions from target sentences was 38.97%.

Table 4.42

Wrap-Up Times (ms) on Target Sentences

	Disconfirming		Control	
Proficiency	*M*	*SD*	*M*	*SD*
Higher	712	189	724	295
Lower	1,317	367	929	290
Total	1,048	426	838	302

Table 4.43

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Sentence Wrap-Up Times

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	1458573.36	1	1458573.36	9.50	.007	.37
Error (P)	2457683.98	16	153605.25			
			Within-participants			
Text Type (T)	313800.10	1	313800.10	13.71	.002	.46
T × P	354945.60	1	354945.60	15.50	.001	.49
Error (T)	366322.97	16	22895.19			
Total	4951325.99	35				

4.3.3.4 Comprehension Questions

A 2 (Text Type: disconfirming, control) × 2 (Proficiency: higher, lower) mixed ANOVA was conducted on the correct answer rates of comprehension questions (see Table 4.44). The results indicated a significant main effect of Text Type, but there were neither a significant main effect of Proficiency, nor an interaction between Text Type and Proficiency (see Table 4.45). As shown in Table 4.44, performance on comprehension questions was poorer for the disconfirming than control passages.

Table 4.44

Mean Accuracy (%) on the Comprehension Questions for Experimental Passages in Experiment 6

	Disconfirming		Control	
	M	*SD*	*M*	*SD*
Higher	87.08	19.55	95.83	7.72
Lower	81.67	18.34	93.33	11.65
Total	84.07	18.53	94.44	9.90

Table 4.45

Summary Table for Two-Way ANOVA of the Effects of Proficiency and Text Type on Mean Comprehension Question Accuracy in Experiment 6

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Between-participants						
Proficiency (P)	0.01	1	0.01	0.52	.482	.03
Error (P)	0.43	16	0.03			
Within-participants						
Text Type (T)	0.09	1	0.09	4.86	.042	.23
T × P	0.00	1	0.02	0.10	.757	.01
Error (T)	0.30	16	0.02			
Total	0.84	35				

4.3.4 Discussion

H5: Japanese EFL learners have some difficulty eliminating disconfirmed predictive inferences from text memory.

The analysis of the SRT revealed that recognition scores of inference sentences were higher for disconfirming than control texts, while there was no significant difference between texts for explicit and inconsistent sentences. These results suggest that disconfirming and control texts only differed in the likelihood of target inference generation: Participants were less likely to generate target inferences when reading control texts compared to disconfirming texts. This supports the claim that the control passages worked as intended in the present experiment.

More importantly, mean recognition score of the inference sentence for the disconfirming passages (M = 2.39, SD = 1.05) was similar to that of Experiment 5 (M = 2.62, SD = 1.20), $t(52) = 0.69$, p = .491, d = 0.20. This score suggests that participants correctly recognized the sentences describing the disconfirmed inferences as not written in the text, but the confidence level was low to relatively low. Thus, the results of the present SRT support the hypothesis that Japanese EFL learners have some difficulty eliminating the memory trace of disconfirmed predictive inferences. The following section discusses what comprehension processes are related to such difficulty based on the results of eye movement data analysis.

RQ4-4: How do the inconsistencies between drawn predictive inferences and the following text affect Japanese EFL learners' text comprehension processes?

The analysis of first-pass reading times demonstrated significantly longer times for disconfirming texts than for control texts. As already noted, disconfirming and control texts only differed in the generation of target inferences. Therefore, the longer first-pass reading time for the disconfirming texts is attributed to the generation of the target inferences during reading the fourth sentences, which made the initial processing of the target sentences more difficult. This suggests that participants noticed inconsistencies between drawn inferences and target sentences during the immediate processing of the sentence, and such awareness resulted in more and longer fixations on target sentences during the first-pass reading. Thus, the results of first-pass reading times decrease the likelihood that the difficulty of revising disconfirmed inferences results from the failure to detect inconsistencies or from insufficient awareness of inconsistencies between drawn inferences and the sentence being processed.

In contrast, the analysis of sentence wrap-up times indicated the significant interaction effect between text type and L2 proficiency, showing wrap-up times significantly longer for disconfirming than control texts only among lower proficiency learners. These results suggest that lower proficiency learners had considerable difficulty integrating the target sentences into developing text representations when reading disconfirming texts. Taken together, the present results suggest that the difficulty of revising predictive inferences is more likely to be related to the additional demand of integrating the disconfirming context into developing text representations.

Although higher proficiency learners showed no significant difference in wrap-up times between text types, this result should be interpreted with care. The results of sentence recognition suggest that the disconfirmed inferences were not completely deleted from text memory, even for higher proficiency learners. Furthermore, the response time analysis in Experiments 4 and 5 suggested that even higher proficiency learners failed to suppress the activation of drawn inferences immediately after reading the disconfirming sentence. Therefore, it is possible that in the present experiment higher proficiency learners also experienced some difficulty integrating disconfirming sentences, although the difficulty was not large enough to be reflected in eye movement data. Nevertheless, the results of wrap-up times at least suggest that lower proficiency learners experienced much larger difficulty than higher proficiency learners in integrating the disconfirming sentences into developing text representations.

However, there was no significant difference in second-pass reading times between text types. As described above, second-pass reading of the target sentences as well as regressions into preceding sentences did not frequently occur in this experiment. This result can be explained in a few ways. One is that, instead of looking back to prior context, readers could rely on their text memory to access prior text information because each passage consisted of only five simple sentences. Another possibility is that readers stayed on the target sentence, having felt that the existing text information was not sufficient to revise the dawn inferences and hoping to find further information in the subsequent context. Furthermore, as noted above, it is possible that eye movements across sentences during L2 reading largely depend on individuals and should be thought of as a personal trait.

On one hand, less frequent second-pass reading and regressions observed in the present experiment supports the validity of the results of the previous sentence-by-sentence experiment (Experiment 5) because participants generally processed the text sequentially, even when the text was presented in its entirety. However, on the other hand, this relates to one of the limitations of the present experiment as noted in Chapter 5.

4.3.5 Conclusion of Experiment 6

Experiment 6 investigated Japanese EFL learners' text comprehension processes when they make predictive inferences and the inferences are disconfirmed by the subsequent context. The results of sentence recognition supported the proposed hypothesis (H5): Consistent with Experiment 5, the participants in the present experiment had some difficulty eliminating memory trace of disconfirmed predictive inferences.

Regarding the RQ4-4, the analysis of eye movements revealed that inconsistencies between learners' predictive inferences and the subsequent context affected some aspects of real-time comprehension processes. First, as indicated by longer first-pass reading times, drawing on predictive inferences that are inconsistent with the following sentences caused extra demand of initial processing of the sentences, regardless of L2 proficiency. This suggests that both higher and lower proficiency learners immediately noticed inconsistencies between drawn inferences and the sentence being processed. Second, as revealed by longer sentence wrap-up times, such inconsistencies also caused difficulty with integrating the subsequent text information into developing text representations, especially among lower proficiency learners. Taken together, the present study suggests that difficulty of revising predictive inferences in memory does not result from insufficient awareness of inconsistencies between drawn inferences and the disconfirming context during reading. Instead, it is more likely to be related to difficult integration of the disconfirming context into developing text representations, which is more demanding for less proficient readers.

4.4 Conclusion of Study 2

Study 2 investigated the revision of predictive inferences in Japanese EFL learners' reading comprehension. Experiments 4 and 5 examined whether Japanese EFL learners suppress the activation of disconfirmed inferences and delete the inferences from long-term text memory. Subsequently, Experiment 6 examined the real-time text comprehension processes of Japanese EFL learners when they encounter context disconfirming predictive inferences.

In Experiment 4, suppression and deletion of predictive inferences were explored using the word-recognition task and a cued recall task, including the consideration of learners' L2 proficiency. The results of these tasks showed that both proficient and less proficient readers failed to suppress the activation of inferences during reading and to delete the inferences from text memory, even after the context disconfirmed the drawn inferences.

The findings of Experiment 4 were reexamined in Experiment 5, using the MJT and the SRT. The results of the response times for the MJT confirmed the previous finding: Both proficient and less proficient readers did not suppress the activation of predictive inferences immediately after the context disconfirming the inferences. In contrast, the results pertaining to the SRT suggested that readers partially deleted the disconfirmed inferences from text memory, although it was still difficult to achieve complete deletion of the inferences.

In Experiment 6, readers' eye movement data were collected to examine real-time comprehension processes when they encountered the context disconfirming the inferences. In addition, readers' text memory was assessed with the same SRT used in Experiment 5. The results of the SRT refined the previous finding that readers had some difficulty eliminating the memory trace of disconfirmed predictive inferences. More importantly, eye movement analysis suggested that both proficient and less proficient readers consciously noticed the inconsistencies between drawn inferences and the disconfirming context during reading;

however, lower proficiency readers experienced more difficulty than higher proficiency readers in integrating the disconfirming context into developing text representations.

In conclusion, the results of Study 2 indicated that it is difficult for Japanese EFL learners to achieve the revision of disconfirmed predictive inferences in reading. Specifically, it is first difficult to suppress the activation of drawn predictive inferences immediately after they are disconfirmed by the context. Second, readers also have difficulty eliminating the disconfirmed inferences from long-term text memory, although partial deletion of the inferences is achieved. Such difficulty in revising predictive inferences is more likely to be related to the additional demand of integrating the disconfirming context with developing text representations, rather than to insufficient awareness of inconsistencies between drawn inferences and disconfirming context. Less proficient readers have more trouble than proficient readers in such an integration process.

Additionally, Study 2 examined the effects of disconfirmation of predictive inferences on Japanese EFL learners' comprehension of explicit text information. However, the obtained results are less clear. The recall task in Experiment 4, the SRT in Experiment 5, and comprehension questions in Experiments 4 to 6 did not produce converging results. This dissociation between experiments is carefully discussed in Chapter 5.

Chapter 5

General Discussion and Conclusion

5.1 Overview of Findings and General Discussion

In this dissertation, two studies examined a total of nine RQs. Study 1 addressed the following five questions regarding the generation of predictive inferences in Japanese EFL learners' reading comprehension:

RQ1-1: Do Japanese EFL readers make predictive inferences differently depending on subtypes of the inferences?

RQ1-2: Do Japanese EFL readers make predictive inferences differently depending on the contextual constraint?

RQ2: Do Japanese EFL readers make predictive inferences differently depending on the amount of cognitive resources available during reading?

RQ3-1: Do Japanese EFL readers strategically make predictive inferences when instructed to anticipate likely outcomes of described events?

RQ3-2: Does strategic processing aimed at predictive inferences affect Japanese EFL readers' comprehension of explicit text information after reading?

First, RQs 1-1 and 1-2 were examined in Experiment 1 by manipulating experimental passages in terms of subtypes of inferences (motivational vs. consequence) and contextual constraint (strong vs. weak). Participants read each passage on the computer screen sentence-by-sentence, and performed the recognition task for an inference-related word immediately after each passage. In addition, they read several passages on a sheet of paper and then engaged in the cued recall task. Correct response times from the recognition task and recall rates of inferential information suggested that motivational inferences were more likely to be made than consequence inferences while reading. Furthermore, the results indicated that motivational inferences were more likely to be drawn when the contextual constraint was strong than weak. In contrast, regardless of the strength of contextual constraint, there was no evidence of the generation of consequence inferences. In sum, the results of Experiment 1 suggested that predictive inferences are most likely to be made when they are motivational (i.e., related to narrative characters' motivation) and strongly constrained by context.

The high likelihood of the generation of motivational inferences that are strongly constrained by the context was confirmed in a subsequent investigation using a lexical decision task in Experiment 2. Similarly, the results of lexical decisions in Experiment 3 provided further evidence that consequence inferences were less likely to be made during normal reading, even when the contextual constraint was strong. Therefore, the converging evidence obtained with different tasks in Experiments 1 to 3 strongly suggested the effects of text characteristics on predictive inference generation in Japanese EFL learners' reading.

Regarding RQ2, Experiment 2 employed the dual-task methodology to investigate the relationship between predictive inference generation and the amount of cognitive resources available during reading. As in Experiment 1, participants read the passages that strongly induced the motivational inferences on the computer screen. Additionally, participants in the dual-task condition were asked to retain a list of words in their memory while reading. Analysis of lexical decision times for inference-related words suggested that predictive inferences were less likely to be made under the dual-task condition (i.e., both low- and high-load conditions) than the single task condition (i.e., zero-load condition), although the effect of the dual-task was not as strong as expected. These results suggest that predictive inference generation is likely to be impaired when the amount of cognitive resources available during reading is reduced. Furthermore, the results of probability judgments of inferences after reading raised the possibility that the reduction of available cognitive resources impaired the immediate or automatic generation of predictive inferences during the course of reading.

Experiment 3 addressed RQ3-1, which concerns whether predictive inference generation is promoted by strategy and task instructions. This experiment used the passages wherein context strongly suggests the generation of consequence inferences. Participants read the passages in either (a) the non-orienting condition, where they were asked to read the passages for accurate comprehension and answer a literal yes/no comprehension question (as in Experiment 1); or (b) the strategic orienting condition, where they were instructed to read the passages for predictions and answer an inference question (i.e., “What will happen next?”). The effects of these strategy and task instructions were examined including consideration of leaners’ L2 proficiency. Although analysis of sentence reading times suggested that learners with both higher and lower proficiency altered their text processing according to instructions, the results of lexical decision times indicated that the facilitation effect of strategy instructions on predictive inference generation was more prominent in higher than lower proficiency readers. Moreover, Experiment 3 investigated RQ3-2, which concerns the effects of strategic processing aimed at predictive inferences on comprehension of explicit text information. The results of the cued recall task demonstrated that explicit text comprehension was not reduced, even when readers’ attention was focused on predictive inference generation by strategy instructions; rather, it likely facilitated comprehension. This trend was consistent between higher and lower proficiency readers.

Study 1 addressed these five RQs (RQs 1-1, 1-2, 2, 3-1, and 3-2) regarding predictive inference generation in Japanese EFL learners’ reading comprehension, and found that it was affected by text characteristics, the amount of cognitive resources available during reading, types of instructions or/and tasks, and learners’ L2 proficiency. Although some L2 studies have found that predictive inferences are made during L2 reading (e.g., Horiba, 1996, 2000; Yoshida, 2003), few studies have examined what specific factors affect the likelihood of this inference generation in detail. Therefore, the findings of Study 1 provide new insight into inference generation during L2 reading.

Study 1 was conducted with short narrative passages consisting of four sentences. Nevertheless, the findings of this study can be partially applied to reading of longer passages. Murray et al. (1993) concluded that passage length is not a critical factor for predictive inference generation. Additionally, Experiment 1 manipulated the local characteristics of the passages: local coherence and contextual constraint of the last sentence of the passage. In general, these text factors are the same regardless of text length. Therefore, the fundamental process of predictive inference generation during EFL reading revealed by Study 1 is assumed to be common to the reading of both short and long passages.

According to the findings of Study 1, predictive inference generation is limited among Japanese EFL readers compared to L1 English readers. For example, Experiment 1 found no evidence of predictive inference generation with consequence inferences, regardless of the strength of contextual constraint. In contrast, previous studies on L1 reading found evidence of consequence inference generation when the contextual constraint was strong (Klin, Guzmán, et al., 1999; Murray et al., 1993). In the discussion of Experiment 1, such limited generation of predictive inferences in EFL reading was attributed to learners' careful reading rather than their limited language proficiency. Indeed, Experiment 3 revealed that proficient learners made consequence inferences during reading when they were encouraged to do so. Thus, one of the possible factors causing the different predictive inference generation between L1 and L2/EFL reading is the reading mode: L2/EFL readers tend to adopt the more careful mode of text processing than L1 readers.

On the other hand, in Experiment 3, less proficient learners provided only minimal evidence of predictive inference generation, even when explicitly instructed to make predictions. Therefore, it is apparent that predictive inference generation during EFL reading is also constrained by readers' language proficiency. In relation, Experiment 2 found that predictive inferences were less likely to be made when readers did not have sufficient cognitive resources available, even though the text strongly induced these inferences. Taken together, it is concluded that predictive inference generation in EFL reading is a more limited and complicated process than in L1 reading as a result of an interaction between text, reader, and task factors.

Regarding such limited generation of predictive inferences in EFL reading, it is possible to state that predictive inferences are generated in EFL reading when both the *necessity* and *sufficiency* of the generation is high (see Figure 5.1). The necessity of inference generation is relevant to maintenance of the local coherence of the passage (i.e., subtypes of inferences), and instructions and/or tasks given to readers. When the inferences are needed to maintain the local coherence of the passage or to achieve a specific task, the necessity of inference generation increases. On the other hand, the sufficiency of inference generation relates to factors such as contextual constraint, the amount of cognitive resources available, and learners' L2 proficiency. When the context strongly constraints a single and specific inference or readers have enough cognitive resources and L2 proficiency to make inferences, the sufficiency of inference generation increases.

For example, Experiment 1 found that consequence inferences are not made during normal reading, even with strong contextual constraint: The sufficiency of inference generation is high (sufficient cognitive resources, strong contextual constraint), but the necessity of inference generation is low (inferences unnecessary for maintaining the local coherence, no strategy instructions or tasks) in this condition. Experiment 2 indicated impairment of inference generation under dual-task conditions, even when the text strongly

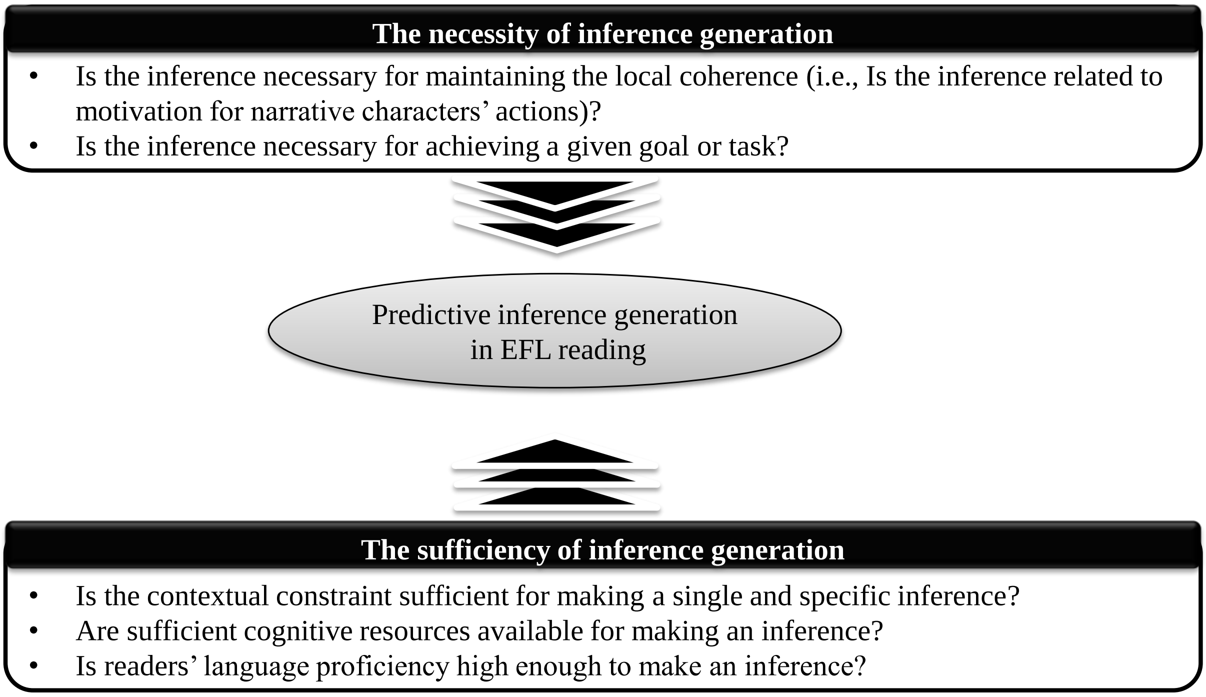

Figure 5.1. The necessity and sufficiency of predictive inference generation in EFL reading.

induced motivational inferences: The necessity is high (inferences necessary for maintaining the local coherence), but the sufficiency is low (insufficient cognitive resources) in this condition. Subsequently, Experiment 3 revealed that predictive inferences were not made among lower proficiency readers even when they were encouraged to do so: The necessity is high (with strategy instructions), but the sufficiency is low (lower L2 proficiency) in this condition. In general, L1 readers make predictive inferences when either the necessity or the sufficiency of the inference generation is high (Iseki, 2006; Klin, Guzmán, et al., 1999; Klin, Murray, et al., 1999; Murray et al., 1993).[24]

[24] Some studies have suggested that low necessity or sufficiency of inference generation might impair the automatic or immediate predictive inference generation during L1 reading (Allbritton, 2004; Linderholm, 2002; Murray & Burke, 2003).

Nevertheless, even when EFL learners make predictive inferences during reading as a result of high necessity and sufficiency of inference generation, they do not always make inferences that are consistent with the following context. If the drawn inference is inconsistent with the subsequent text, it is necessary to revise the drawn inferences and construct accurate text comprehension. Study 2 examined the revision of predictive inferences in Japanese EFL leaners' reading comprehension and addressed the following four questions:

RQ4-1: Do Japanese EFL readers suppress the activation of predictive inferences when the subsequent context disconfirms the inferences?

RQ4-2: Do Japanese EFL readers delete the disconfirmed predictive inferences from long-term text memory?

RQ4-3: How does the disconfirmation of predictive inferences influence Japanese EFL readers' comprehension of explicit text information?

RQ4-4: How do the inconsistencies between drawn predictive inferences and the following text affect Japanese EFL learners' text comprehension processes?

RQs 4-1 and 4-2 were examined in Experiment 4 by comparing task performance when reading predictive and disconfirming texts. Predictive texts induce the motivational inferences strongly constrained by the context (i.e., both the necessity and sufficiency of inference generation is high), whereas each disconfirming text included an additional sentence that disconfirms the inference drawn from prior context. As in Experiment 1, participants performed the recognition task for inference-related words immediately after reading each passage, and engaged in the cued recall task after reading all of the passages. The analysis of correct recognition times and inference recall rates did not show significant differences between predictive and disconfirming text conditions, regardless of learners' L2 proficiency level. These results suggested that the activation of predictive inferences was not suppressed immediately after the disconfirming context, and that memory of the disconfirmed inferences were not deleted.

However, some methodological issues remained in Experiment 4. Therefore, the suppression and deletion of disconfirmed predictive inferences were reexamined in Experiment 5 with different tasks. In this experiment, participants performed the MJT for the target sentences instead of the word-recognition task, and engaged in the SRT instead of the cued recall task. Consistent with Experiment 4, the analysis of correct judgment times indicated that the activation of predictive inferences was maintained, even after the subsequent context disconfirmed the inferences. The analysis of sentence recognition suggested that it was difficult to achieve complete deletion of disconfirmed inferences from memory, although the inferences were partially deleted. These trends were consistent between higher and lower proficiency readers. Therefore, the results of Experiments 4 and 5 revealed that when leaners make a predictive inference and the subsequent context disconfirms it, they experience difficulty suppressing the activation of the inference immediately after the disconfirming context and deleting the disconfirmed inference from long-term text memory.

Experiment 6 addressed RQ5, which concerns the text comprehension process when readers encounter the context disconfirming predictive inferences. In the experiment, an eye tracking methodology was adopted to examine what comprehension processes are related to difficulty revising predictive inferences, as well as to overcome the methodological limitations of Experiments 4 and 5 (e.g., sentence-by-sentence presentation of experimental passages). In addition, learners' text memory was investigated by the same SRT used in Experiment 5 to refine the previous finding. The results of sentence recognition confirmed that Japanese EFL readers had some difficulty eliminating the memory trace of disconfirmed predictive inferences after reading, regardless of L2 proficiency. The analysis of eye movements on the disconfirming context suggested that both proficient and less proficient readers noticed the inconsistencies between drawn inferences and the disconfirming context during the immediate processing of the context. In contrast, the results further revealed that lower proficiency readers experienced more difficulty than higher proficiency readers in integrating the disconfirming context into developing text representations.

In sum, the major finding of Study 2 is that when Japanese EFL readers strongly activate a predictive inference and it is then disconfirmed by the following context, they have some difficulty revising the inference. Specifically, they are likely to fail in suppressing the activation of the inference immediately after the disconfirming context. In addition, although partial deletion of the inference is achieved, it is difficult to completely delete the disconfirmed inference from long-term text memory. Such difficulty revising predictive inferences is assumed to be related to the difficult integration of the disconfirming context into developing text representations during text processing, which is more demanding for less proficient than proficient EFL readers.

Study 2 also suggested some possible factors affecting the difficulty of revising predictive inferences. The most plausible factor is the amount and quality of the disconfirming context. Experiments 4 and 5 pointed out the possibility that the single disconfirming sentence was not enough to fully suppress and delete the disconfirmed inferences. Furthermore, eye movement data in Experiment 6 revealed that readers stayed with the disconfirming sentence instead of looking back to prior contexts, suggesting that they felt that the existing text information was not sufficient to revise the inferences. Another possible factor is the characteristics of induced inferences. Because experimental passages strongly induced inferences related to narrative characters' motivation, the activation strength might be quite high or the revision of predictive inferences might not always be necessary. In the discussion of Experiment 5, the effect of text presentation mode on the difficulty of revising predictive inferences was also mentioned. However, this is unlikely because even when the passage was presented in its entirety in Experiment 6, participants made few regressions to prior sentences during reading (i.e., read the text sequentially sentence-by-sentence) and continued to have difficulty processing the disconfirming sentence.

Experiments 4 and 5 did not show a significant effect of leaners' L2 reading proficiency on revision of disconfirmed inferences. Based on the discussion of Experiment 5, it is likely that the aforementioned characteristics of experimental passages so strongly affected the difficulty of suppression and deletion of the inferences that the effect of L2 proficiency was less clear. Nevertheless, the results of response times in Experiments 4 and 5 raise the possibility that L2 proficiency had some impacts on the suppression of predictive inferences. In these experiments, higher proficiency readers showed more noticeable differences in activation scores between the text conditions (i.e., predictive and disconfirming texts) compared to lower proficiency readers, although the difference did not approach statistical significance. For example, descriptively speaking, higher proficiency readers in Experiment 4 showed higher activation scores in the disconfirming than predictive text condition, whereas those in Experiment 5 showed higher activation scores in the predictive than disconfirming text condition. Although these results revealed opposite patterns, they may be explained by considering the unit of inferences. Given that different units of target probes were used in the two experiments (i.e., a single word in Experiment 4 vs. a single sentence in Experiment 5), the disconfirmation of predictive inferences might enhance the activation of an inference as a single concept (e.g., *hit*), but it might suppress the activation of the inference as a set of concepts constituting a proposition (e.g., *The boy hits the ball*).[25]

[25] The pattern of the results of Experiment 4 is similar to that of Potts et al. (1988), who used the lexical decision and naming tasks for target words, whereas the results of Experiment 5 is consistent with Iseki

Although this interpretation is merely speculative and needs to be verified by further investigations, it is at least possible to state that higher proficiency readers were more sensitive to the manipulation of experimental passages than lower proficiency readers.

Similarly, the results of sentence recognition in Experiments 5 and 6 raise the possibility that L2 proficiency affected the deletion of predictive inferences. In these experiments, the mean recognition scores of the inference sentences for disconfirming texts were lower in higher than lower proficiency readers, although the difference was not statistically significant. This suggests that higher proficiency readers were more confident than lower proficiency readers in recognizing the inference sentences as not written in the disconfirming texts; in other words, higher proficiency readers were more successful in deleting the disconfirmed inferences from text memory than lower proficiency readers.

Unlike the tasks used in Experiments 4 and 5, eye movement analysis in Experiment 6 showed some distinct differences in the processing of disconfirmed inferences between higher and lower proficiency readers (e.g., longer sentence wrap-up times for less proficient than more proficient readers). This might be the result of adopting a different methodology from that used in previous experiments (i.e., eye fixation measures and the presentation of entire passages.)

In addition to the suppression and deletion of inferences, RQ4-3 in Experiment 4 examined the effects of disconfirmation of inferences on explicit text comprehension. The amount of recalled text information was not significantly different between predictive and disconfirming text conditions; however, descriptive statistics and qualitative observation of recall protocols raised the possibility that the disconfirmation of inferences reduced explicit text comprehension and made comprehension less coherent. Consistent with this idea, in Experiment 6, the correct answer rates for comprehension questions were significantly lower

(2006), who used the MJT for the target sentences.

in the disconfirming than the control text conditions. On the other hand, the results of sentence recognition in Experiments 5 and 6 revealed that disconfirmation of inferences did not affect the accuracy of recognizing the sentences describing explicit text information. Similarly, in Experiments 4 and 5, there was no significant difference in the correct answer rates for comprehension questions between the text conditions. The dissociation between these results can be explained by two possibilities.

One possibility is that the effects of disconfirmation of inferences on explicit text comprehension vary depending on the measure of comprehension. Specifically, the effects may be clearer when the quality (and perhaps quantity) of recalled text information is assessed (Experiment 4). On the other hand, the effects may become smaller when the task requires less cognitive effort, such as recognizing the explicit text information described in participants' L1 (Experiments 5 and 6).

Another possibility is that the effects of disconfirmation of inferences vary depending on what condition is compared to the disconfirming text condition. In Experiments 4 and 5, the performance in the disconfirming text condition was compared to that of the predictive text condition, and no significant differences in comprehension question answer rates were shown between the conditions. On the other hand, in Experiment 6, the performance in the disconfirming text condition was compared to that of the control text condition, and a significant difference in comprehension question answer rates was shown between the conditions. The control passages in Experiment 6 were created by rewriting the disconfirming passages so that they did not induce the target predictive inferences and include any disconfirmation. Consequently, the participants in Experiment 6 processed the control passages more easily than disconfirming passages, as reflected by fewer and shorter fixations for the control texts. Therefore, the control and disconfirming passages greatly differed in terms of textual coherence and processing demands, which might result in the impaired explicit text comprehension for the disconfirming passages.

Unfortunately, because the results pertaining to explicit text comprehension are mixed and complicated between experiments, it is difficult to draw a conclusion about the effects of disconfirmation of predictive inferences on explicit text comprehension in this dissertation. Nevertheless, it is highly possible that the disconfirmation of drawn predictive inferences have some negative effects on comprehension of explicit text information depending on the situation.

5.2 Limitations of the Present Studies and Suggestion for Future Research

While the findings of the present studies offer new insight into understanding predictive inferences in EFL reading, some limitations should be noted, as well as promising directions for future research. First, general limitations that are common between all of the experiments are described below. Then, some limitations regarding the individual experiments are noted. In addition, possible directions for future research are discussed throughout.

General limitations

First, the present studies are limited by participant characteristics, such as L2 proficiency (i.e., intermediate level) and L1 background (i.e., Japanese). In other words, it remains uncertain how predictive inferences are made and revised among advanced or novice level Japanese EFL readers. Similarly, predictive inferences may not always be made and revised similarly between Japanese EFL readers and EFL readers who have an L1 that is an alphabetic language. Therefore, generalizing the findings of the present studies requires addressing these issues. Furthermore, it will be valuable to compare the generation and revision of predictive inferences between L1 and L2 reading (e.g., Japanese as L1 vs. English as L2 among Japanese EFL learners, or native speakers of English vs. Japanese learners of English) within a single experiment. Comparing L1 and L2 reading has often been conducted in previous studies on discourse processing (e.g., Horiba, 1996, 2000; Morishima, 2013), and it allows for the clarification of difficulty inherent in L2 reading.

Second, the present findings are limited by the methods and tasks used. Although some of the present RQs were investigated by implementing different methods (e.g., RQs1-1 and 4-1), the present findings still need to be confirmed with tasks or measures that differ from those used in the present studies. For example, Study 1 assessed inference generation with a word-recognition and lexical decision tasks, but some researchers have claimed that these tasks may be affected by context checking (Keenan et al., 1990). Therefore, it is important to refine the present findings regarding predictive inference generation by combining the tasks used here with other tasks. For example, Magliano and Graesser (1991) suggested the use of a three-pronged method for investigating inference generation that includes: (a) collection of on-line response time data, (b) collection of think-aloud protocols, and (c) theories of discourse processing (see also Graesser et al., 1994).

Third, the present findings might be constrained by the characteristics of experimental passages, such as text length, difficulty, and familiarity of the content. For example, because response time research requires as many trials as possible, the experimental passages were several short narratives. Although passage length is assumed to not be a critical factor for predictive inference generation (Murray et al., 1993), it is interesting to investigate how predictive inferences are made and revised when reading comparatively long narrative passages such as those adopted from EFL reading textbooks. This approach will be both theoretically and pedagogically important to address in future research on EFL reading.

Limitations of individual experiments

There are several limitations regarding the individual experiments. First, Experiments 1 and 2 did not include consideration of reader factors. Therefore, it remains unclear how reader factors (e.g., L2 proficiency or WM capacity) interact with other factors (e.g., text factors) in making predictive inferences in EFL reading. Future studies should examine an interaction effect of reader and other factors on predictive inference generation in EFL reading.

In particular, individual differences in WM capacity may be an important factor in Experiment 2. This is because it is difficult to state that the size of the cognitive load given was totally equal between the participants without considering individual differences in WM. For example, the demand of a high load might not be actually high for readers with high WM capacity. Given this possibility, the results of Experiment 2 could be extended by including a direct measure of participants' WM capacity.

Although Experiments 3 to 6 included L2 proficiency as a critical reader factor, it will be valuable to conduct similar experiments involving participants with more L2 proficiency variations. Experiment 3 indicated that strategy instructions aimed at predictive inference generation did not reduce comprehension of explicit text comprehension for participants with intermediate-level English proficiency. However, Koda (2005) suggested that instructions focusing on higher level text processing (including making inferences) could negatively affect reading comprehension for novice level L2 learners. Given this suggestion, it is important to examine how strategy instructions aimed at predictive inferences affect novice level EFL learners' reading comprehension.

Additionally, in Experiments 4 and 5, there was no significant effect of L2 proficiency on the revision of predictive inferences. Although the previous section suggested some possible effects of L2 proficiency, these effects were less clear, perhaps due to the limited range of participants' L2 proficiency. Thus, it is also necessary to examine the revision of inferences among EFL readers with a broader range of L2 proficiency. Although some results of Experiment 6 showed distinctive differences between higher and lower proficiency readers, higher proficiency readers demonstrated dissociation between the results of eye movement measures and sentence recognition. Therefore, the effect of L2 proficiency on the processing of the disconfirming context and memory of inferences disconfirmed by the context also requires further investigations.

In Experiment 2, a word-recall task was used to tax readers' cognitive resources based on Shears et al. (2007). However, the use of other tasks or methods may allow us to clarify the exact nature of the cognitive resources related to predictive inference generation. Fincher-Kiefer and D'Agostino (2004) suggested that making predictive inferences in reading requires in particular visuospatial resources. In addition, Nakanishi and Yokokawa (2011) suggested that syntactic processing places a heavy load on WM resources among Japanese EFL readers relative to other types of processing (e.g., semantic or pragmatic processing). Therefore, it seems interesting to investigate how visuospatial processing or syntactic processing difficulty relates to the generation of predictive inferences among Japanese EFL readers. Given that Experiment 2 indicated only a small effect of verbal WM demands on predictive inference generation, these investigations are quite necessary. The difficulty of syntactic processing may be varied by manipulating the syntactic structure of experimental passages (e.g., type of relative clauses). Such manipulation would allow the relationship between cognitive demands and predictive inference generation to be examined under more natural reading conditions, unlike Experiment 2 where participants were asked to hold a list of words in memory during reading.

Regarding the response task used in Experiments 2 and 3, it is necessary to carefully consider the use of lexical decisions as a measure of inference generation. Although probe characteristics were carefully controlled and the participants performed better on the lexical decision task in these experiments, further research is also required on whether the task adequately reflects inference activation by EFL readers, especially in light of the issue of mental translation (Kroll & Stewart, 1994). The MJT used in Experiment 5 may have a similar potential problem.

In Experiment 3, the orienting condition was a within-participants factor because the focus of this experiment was on how individuals alter their reading process according to instructions. Participants always read passages in the non-orienting condition prior to the strategic orienting condition so that the strategic orienting instructions did not influence text processing in the non-orienting condition. However, it is difficult to fully rule out the influence of the preceding non-orienting condition trials on the performance in the strategic orienting condition. Thus, an additional experiment should be conducted that includes orienting condition as a between-participants factor or utilizes a control group that performs twice in the non-orienting condition.

One common limitation between Experiments 4, 5, and 6 is that the context disconfirming target inferences consisted of a single sentence. As discussed earlier, the disconfirming information described by a single sentence might not be sufficient for readers to fully revise the inferences. Therefore, future studies need to clarify the effects of contextual factors (e.g., quantity and quality of disconfirming information) on the revision of predictive inferences. More specifically, the revision of inferences might be facilitated if readers receive more disconfirming information in several sentences, or if the subsequent context clearly disconfirms the drawn inferences and induces alternative inferences. It would also be interesting to examine how readers revise their inferential comprehension sentence-by-sentence using a method similar to that used by Ushiro (2010). As Tapiero (2007) noted, in some cases, suppressed information is reactivated during reading and then encoded into long-term text memory. Such flexible and dynamic comprehension processes in EFL reading may be investigated by manipulating the subsequent context after drawn inferences are disconfirmed.

The limitation regarding the amount of disconfirming context relates to an important methodological limitation in Experiment 6. Specifically, because there were no successive sentences that followed the disconfirming sentence in this experiment, second-pass reading times on the target sentence were confined to forward fixations from prior sentences. Thus, regressive or backward fixations from successive sentences to target sentences were not examined. Therefore, future research should examine how frequently and for how long learners regress to the sentence disconfirming the inferences from the subsequent context.

In addition to the context factors, it is necessary to examine the effects of other factors that possibly affect the difficulty of revising predictive inferences. Such factors should include the characteristics of inferences and types of instructions and/or tasks. Future research that examines the effects of these factors will be valuable because it provides more specific implications for teacher support in facilitating learners' revision of disconfirmed predictive inferences.

5.3 Pedagogical Implications

The present findings have a number of implications for EFL reading instruction. These implications concern instructions focusing on the generation of predictive inferences and development of the ability to revise inferential comprehension in reading.

Reading instruction focusing on the generation of predictive inferences

As mentioned in the review of related literature, previous studies noted the significant role of predictive inferences in reading, such as enabling the smooth processing of the subsequent context and promoting the construction of situation models. In other words, predictive inference generation contributes to achieving reading fluency and enriched text comprehension. However, as revealed in Study 1, teachers should know that EFL students may not always make predictive inferences during reading according to text, reader, and task factors. For example, as shown in Experiment 1, predictive inferences that are not related to narrative characters' motivation are less likely to be drawn during normal reading, even though the future event is highly predictable. Thus, teachers need to consider not only how their students can achieve accurate comprehension of text meaning, but also how students are encouraged to engage in the generation of predictive inferences during reading.

When teachers consider how to enhance students' predictive generation in reading, it is first important to pay attention to text characteristics. When the text has low WM demands (i.e., less difficult for students to comprehend) and the context strongly induces the predictive inference that is related to narrative characters' motivation, instructional interventions by teachers may be unnecessary because students are more likely to spontaneously make these inferences. However, when the context strongly constrains the inference but it is less related to characters' motivation, it is recommended that teachers provide pre-reading instructions and questions to direct students' attention toward predictive inferences in the same way as in Experiment 3. That said, according to the findings of Experiment 3, such instructions and questions might only be effective for on-line generation of predictive inferences in higher proficiency readers. Nevertheless, in contrast to giving an instruction to simply comprehend the passage and asking yes/no text comprehension questions, strategy instructions or questions aimed at predictive inferences will encourage both higher and lower proficiency learners' active engagement in and better comprehension of the text, as indicated by sentence reading time and recall protocol analyses in Experiment 3. In addition, strategy or task instructions aimed at predictive inference generation can be valuable for every student in that they "help the students experience different modes of text processing, become more aware of their own processes, and develop flexible strategic reading proficiency" (Horiba, 2013, p. 118).

However, Koda (2005) pointed out that when students have difficulty with lower level reading processing, instructions focusing on higher level text processing can be demanding and frustrating for them. Therefore, teachers should carefully consider the balance between their students' English proficiency level and text difficulty when instructing them to generate predictions while reading. If teachers want less proficient students to make predictions, it might be useful to simplify the text or gloss over unknown words to reduce the demands on lower level text processing.

In relation, the results of Experiment 2, which suggested some impairment of the inferences when the amount of cognitive resources available during reading is reduced, emphasize the importance of improving learners' lower level processing abilities (e.g., word recognition, syntactic analysis). Effective or automatized lower level processing allows more cognitive resources to be used for generating predictive inferences. Therefore, training in lower level processing (e.g., extensive reading or reading aloud) not only improves processing itself, but also becomes a significant step toward actively generating predictive inferences in reading.

Developing learners' ability to revise inferential comprehension

Successful text processing requires readers to allocate attentional resources flexibly and efficiently to comprehend the text, including the suppression of irrelevant text information (Gernsbacher et al., 2004; Horiba, 2000). Similarly, flexible revision of text comprehension plays an important role in reading (Ushiro, 2010).

However, the results of Study 2 suggest that when learners make strong predictions that are immediately disconfirmed in the following context, they may have trouble suppressing the activation of the predictions and eliminating the memory trace of the predictions. Furthermore, learners' comprehension of explicit text information may sometimes be reduced by the disconfirmation of the inferences. In sum, it is possible that students construct less accurate text representations when they make incorrect predictive inferences during reading. Therefore, teachers should focus not only on their students' active generation of predictive inferences in reading, but also on how to help them successfully revise the inferences when they are incorrect.

Eye tracking data obtained in Experiment 6 suggests that difficulty in revising disconfirmed predictive inferences is more likely to be related to the integration of the disconfirming context with current text comprehension, which is especially demanding for less proficient readers. Therefore, it is important to develop the ability to flexibly integrate unexpected text information with current comprehension during EFL reading, especially among less proficient students. Although it is difficult to propose specific instructions from findings of the present study, it will at least be helpful to give students some questions concerning the predictions (e.g., *What happened to the batter?* for the sample passage in Table 4.1) and ask them to provide reasons for their answers, rather than simply asking them to translate the disconfirming context into Japanese. This may facilitate a careful reanalysis of text information and comprehension monitoring. In this way, teachers can provide the students with good opportunities to flexibly revise their existing comprehension as the text unfolds. The findings of Study 2 emphasize the importance of considering such an approach to develop learners' flexible comprehension processes, and suggest these comprehension processes as one aspect that characterizes proficient and effective reading.

Additionally, the findings of Study 2 are suggestive for EFL text and material writers. Reading stories with unpredicted events or surprising endings can be enjoyable for students.

However, Study 2 suggests that when learners read such stories without sufficient information that encourages revision of comprehension, they experience difficulty revising their comprehension, which might lead to the construction of less accurate text representations. Therefore, writers should include quantitatively and qualitatively sufficient information in these texts so that students can fully comprehend and enjoy reading the stories. If these stories are carefully designed, they will be not only enjoyable, but also useful in developing students' flexible text comprehension processes.

5.4 Concluding Remarks

To date, a number of empirical studies have investigated inference generation in L2 reading (e.g., Collins & Tajika, 1997; Horiba, 1996, 2000; Muramoto, 2000; Shimizu, 2012; Yoshida, 2003). However, few studies have devoted particular attention to predictive inference generation in Japanese EFL learners' reading comprehension, despite its benefits for reading comprehension. Additionally, few have investigated whether and how these inferences are revised when the subsequent context disconfirms the inferences.

Study 1 of this dissertation examined predictive inference generation in Japanese EFL learners' reading comprehension by combining on-line data (e.g., response times to target probes) and off-line data (e.g., recall productions). The results of this study revealed the conditions in which Japanese EFL learners are more or less likely to make predictive inferences while reading.

Study 2 investigated the revision of predictive inferences disconfirmed by the subsequent context based on on-line data (e.g., response times to target probes, eye movement measures) and off-line data (e.g., recall productions, recognition performance). The results of this study indicated that Japanese EFL learners have some difficulty revising the disconfirmed inferences, and that less proficient learners in particular have difficulty integrating the context disconfirming the inferences with current comprehension during reading.

These findings contribute to an understanding of how Japanese EFL learners make and revise predictive inferences in reading. The author hopes that the current research will assist with future empirical and educational research on EFL learners' reading comprehension.

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1061–1070. doi:10.1037/0278-7393.19.5.1061

Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, *13*, 219–223. doi:10.2307/3586211

Alderson, J. C. (1980). Native and nonnative speaker performance on cloze tests. *Language Learning*, *30*, 59–76. doi:10.1111/j.1467-1770.1980.tb00151.x

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes*, *38*, 309–322. doi:10.1207/ s15326950dp3803_2

Calvo, M. G., & Castillo, M. D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes*, *22*, 57–78. doi: 10.1080/01638539609544966

Calvo, M. G., & Castillo, M. D. (1998). Predictive inferences take time to develop. *Psychological Research*, *61*, 249–260. doi:10.1007/s004260050029

Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-line predictive inferences in reading: Processing time during versus after the priming context. *Memory & Cognition*, *27*, 834–843. doi:10.3758/BF03198536

Calvo, M. G., Castillo, M. D., & Schmalhofer, F. (2006). Strategic influence on the time course of predictive inferences in reading. *Memory & Cognition*, *34*, 68–77. doi:10.3758/BF03193387

Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research*, *65*, 158–169. doi: 10.1007/s004260000050

Campion, N., Martins, D., & Wilhelm, A. (2009). Contradictions and predictions: Two sources of uncertainty that raise the cognitive interest of readers. *Discourse Processes, 46*, 341–368. doi:10.1080/01638530802629125

Campion, N., & Rossi, J. P. (2001). Associative and causal constraints in the process of generating predictive inferences. *Discourse Processes*, *31*, 263–291. doi:10.1207 /S15326950dp31-3_3

Carrell, P. L. (1989). Metacognitive awareness and second language reading. *The Modern Language Journal*, *73*, 121–134.

Carrell, P. L. (1992). Awareness of text structure: Effects on recall. *Language Learning*, *42,* 1–20. doi:10.1111/j.1467-1770.1992.tb00698.x

Casteel, M. A. (2007). Contextual support and predictive inferences: What do readers generate and keep available for use? *Discourse Processes*, *44,* 51–72. doi: 10.1080/01638530701285572

Chikalanga, I. (1992). A suggested taxonomy of inferences for the reading teacher. *Reading in a Foreign Language*, *8*, 697–709.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, S., & Tajika, H. (1996). Do EFL learners make instrumental inferences when reading? Some evidence from implicit memory tests. *JALT (Japan Association for Language Teaching) Journal*, *18*, 27–39.

Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language*, *44*, 220–234. doi:10.1006/jmla.2000.2744

Council of Europe. (2001). *Common European Framework of reference for languages: Learning, teaching, assessment*. New York, NY: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213–238. doi:10.2307/3587951

Day, R., & Park, J. (2005). Developing reading comprehension questions. *Reading in a Foreign Language, 17*, 60–73.

de Groot, A. M. B., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47*, 91–124. doi: 10.1006/jmla.2001.2840

Dunlea, J. (2010). The relationship between the EIKEN Test and CEFR: A report of research project. Retrieved from http://www.eiken.or.jp/eiken/group/result/pdf/report_02.pdf

Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*, 149–166. doi:10.1017/S0267 19051000005X

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*, 367–413. doi:10.1111/j.1467-9922.2010.00613.x

Educational Testing Service (ETS). (2012). *Official guide to the TOEFL test* (4th ed.): New York, NY: McGraw-Hill.

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes*, *16*, 99–124. doi:10.1080/01638539309544831

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 981–995. doi:10.1037/0278-7393.21.4.981

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, *22*, 225–246. doi:10.1080/01638539609544974

Fincher-Kiefer, R., & D'Agostino, P. R. (2004). The role of visuospatial resources in generating predictive and bridging inferences. *Discourse Processes*, *37*, 205–224.

doi:10.1207/s15326950dp3703_2

Freedle, R., & Halle, G. (1979). Acquisition of new comprehension schemata for expository prose by transfer of a narrative schema. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 121–135). New Jersey, NJ: Ablex.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, *23*, 265–304. doi:10.1080/01638539709544994

Gernsbacher, M. A., Robertson, R. R. W., Palladino, P., & Werner, N. K. (2004). Managing mental representations during narrative comprehension. *Discourse Processes*, *37*, 145–164. doi:10.1207/s15326950dp3702_4

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York, NY: Cambridge University Press.

Graesser, A. C. (1981). *Prose comprehension beyond the word*. New York, NY: Springer-Verlag.

Graesser, A. C., & Kreuz, R. J. (1993). A theory of inference generation during text comprehension. *Discourse Processes*, *16*, 145–160. doi: 10.1080/01638539309544833

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395. doi:10.1037/0033-295X.101.3.371

Graesser, A. C., Wiemer-Hastings, P., & Wiemer-Hastings, K. (2001). Constructing inferences and relations during text comprehension. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 249–271). Amsterdam, Netherlands: John Benjamins.

Guéraud, S., Harmon, M. E., & Peracchi, K. A. (2005). Updating situation models: The memory-based contribution. *Discourse Processes*, *39*, 243–263. doi:10.1080/0163853X.2005.9651682

Gygax, P., Tapiero, I., & Carruzzo, E. (2007). Emotion inferences during reading comprehension: What evidence can the self-pace reading paradigm provide? *Discourse Processes*, *44*, 33–50. doi:10.1080/01638530701285564

Horiba, Y. (1990). Narrative comprehension processes: A study of native and non-native readers of Japanese. *The Modern Language Journal*, *74*, 188–202.

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

Horiba, Y. (2013). Task-induced strategic processing in L2 text comprehension. *Reading in a Foreign Language*, *25*, 98–125.

Horiba, Y., van den Broek, P., & Fletcher, C. R. (1993). Second language readers' memory for narrative texts: Evidence for structure-preserving top−down processing. *Language Learning*, *43*, 345–372. doi:10.1111/j.1467-1770.1993.tb00618

Hyönä, J., Lorch, R. F., Jr., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam, Netherlands: Elsevier Science.

Ikeno, O. (1996). The effects of text-structure-guiding questions on comprehension of texts with varying linguistic difficulties. *JACET (The Japan Association of College English Teachers) Bulletin*, *27*, 51–68.

Iseki, R. (2006). Text rikai ni okeru on-line suiron seisei no kiteiin: Seigousei to access kanousei no hikaku [Which regulates on-line inference generation in text comprehension, coherence or accessibility?]. *Cognitive Studies, 13*, 205–224.

Iseki, R., & Kawasaki, E. (2006). Monogatari to setsumeibun no joukyou model ha donoyouni kotonaruka- itsutsu no joukyouteki jigen ni motozuku hikaku [How do situation models differ in narrative and expository text? A comparison based on five situational dimensions]. *Japanese Association of Educational Psychology, 54*, 464–475.

JACET Committee of Revising the JACET Basic Words (Ed.). (2003). *JACET list of 8000 basic words*. Tokyo, Japan: Author.

Jiang, N. (2012). *Conducting reaction time research in second language studies.* New York, NY: Routledge.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1420–1436. doi:10.1037/0278-7393.20.6.1420.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354. doi:10.1037/0033-295X.87.4.329.

Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eye tracking, and recalls. *Discourse Processes, 40*, 239–257. doi:10.1207/s15326950dp4003_4

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2002). Perspective effects on on-line text processing. *Discourse Processes*, *33*, 159–173. doi:10.1207/S15326950DP3302_03

Kadota, S., & Kuramoto, A. (2008). The process of comprehending GP sentences by Japanese EFL students: A psycholinguistic analysis of eye-movement data. *Japan Society for*

Speech Sciences, *8*, 19–43.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language*, *32*, 446–463. doi: 10.1006/jmla.1993.1024

Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990). Which elaborative inferences are drawn during reading? A question of methodologies. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 377–402). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *39*, 854–865. doi:10.1037/a0029468.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182. doi: 10.1037/0033-295X.95.2.163

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

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394. doi:10.1037/0033-295X.85.5.363

Klin, C. M., Guzmán, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language*, *40*, 593–604. doi:10.1006/jmla.1998.2628

Klin, C. M., Murray, J. D., Levine, W. H., & Guzmán, A. E. (1999). Forward inferences: From activation to long-term memory. *Discourse Processes*, *27*, 241–260. doi: 10.1080/01638539909545062

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. New

York, NY: Cambridge University Press.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174. doi:10.1006/jmla.1994.1008

Lassonde, K. A., & O'Brien, E. J. (2009). Contextual specificity in the activation of predictive inferences. *Discourse Processes*, *46*, 426–438. doi:10.1080/01638530902959620

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, *25*, 259–284. doi:10.1080/01638539809545028

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, *34*, 259–280. doi: 10.1006/jmla.1994.1008

Linderholm, T., Gernsbacher, M. A., van den Broke, P., Neninde, L., Robertson, R. R. W., & Sundermier, B. (2004). Suppression of story character goals during reading. *Discourse Processes*, *37*, 67–78. doi:10.1207/s15326950dp3701_4

Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1424–1429. doi:10.1037/0278-7393.27.6.1424

Long, D. L., & Golding, J. M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes*, *16*, 55–73. doi: 10.1080/01638539309544829

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language*, *31*, 634–647. doi: 10.1016/0749-596X(92)90032-S

Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inference. *Discourse Processes, 16*, 35–53. doi:10.1080/01638539309544828

Magliano, J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, *20*, 193–232. doi:10.1016/0304-422X(91)90007-C

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615–629. doi: 10.1037/0022-0663.91.4.615

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151. doi:10.1016/0010-0285(77)90006-8

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19*, 668–682. doi:10.1016/S0022-5371(80)90355-2

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 82–91. doi: 10.1037/0278-7393.12.1.82

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466. doi:10.1037/0033-295X.99.3.440

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

Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Processes*, *50*, 577–597. doi:10.1080/0163853X.2013.846964.

Motyka Joss, L. S. (2010). *Predictive inference generation in the cerebral hemispheres: An investigation of time course and reader goals* (Doctoral dissertation). Retrieved from College of Liberal Arts & Social Sciences Theses and Dissertations (Paper 23).

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

Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes*, *35*, 81–102. doi:10.1207/S15326950DP3502_1

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language*, *32*, 464–473. doi: 10.1006/jmla.1993.1025

Nahatame, S. (2010). Effects of predictive inferences on Japanese EFL learners' reading comprehension: Focusing on goal information in narrative. *KATE (Kantou-Koushinetsu Association of Teachers of English) Bulletin*, *24*, 49–60.

Nahatame, S. (2011). How predictive inference generation influences Japanese EFL learners' reading: Focusing on comprehension, sentence processing, and the timing of making inferences. *ARELE (Annual Review of English Language Education in Japan)*, *22*, 79–94.

Nahatame, S. (2012). Latent Semantic Analysis (LSA) ni yoru kusyo hojyugata dokkai test no kaimei [Investigating gap-filling reading tests using Latent Semantic Analysis]. *STEP BULLETIN*, *24*, 42–58.

Nakanishi, H., & Yokokawa, H. (2011). Determinant processing factors of recall performance in reading span tests: An empirical study of Japanese EFL learners. *JACET (The Japan Association of College English Teachers) Journal*, *53*, 93–108.

Nishiyama, M., & Kadota, S. (2004). Gaikokugo to shite no Eigo no yomi ni okeru gankyu undou: Nihonjin eigo gakushusya ni taisuru yobiteki kenkyu [Measuring basic data on the eye movements in reading English: A preliminary study for Japanese EFL learners]. *Proceedings of the 10th Annual Meeting of the Association for Natural Language Processing*.

Obunsha. (2009). Eiken jyun nikyu zen mondai-shu [The collection of previous questions for the pre-second grade Eiken test]. Tokyo: Author.

Obunsha. (2010a). Eiken ni kyu zen mondai-shu [The collection of previous questions for the second grade Eiken test]. Tokyo: Author.

Obunsha. (2010b). Eiken jyun ikkyu zen mondai-shu [The collection of previous questions for the pre-first grade Eiken test]. Tokyo: Author.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1200–1210. doi:10.1037/0278-7393.24.5.1200

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal-teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117–175. doi: 10.1207/s1532690xci0102_1

Paris, S. G., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, *8*, 217–227. doi:10.1016/0010-0285 (76)90024-4

Porte, G. (Ed.). (2012). *Replication research in applied linguistics.* New York, NY: Cambridge University Press.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inference: Lexical decision versus naming. *Journal of Memory and Language*, *27*, 399–415. doi:10.1016/0749-596X(88)90064-2

Poynor, D. V., & Morris, R. K. (2003). Inferred goals in narratives: Evidence from self-paced reading, recall, eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 3–9. doi:10.1037/0278-7393.29.1.3

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition, 35*, 2019–2032. doi:10.3758/BF03192934

Rapp, D. N., & Kendeou, P. (2009). Noticing and revising discrepancies as texts unfold. *Discourse Processes*, *46*, 1–24. doi:10.1080/01638530802629141

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506. doi: 10.1080/17470210902816461

Rayner, K., Chace, K., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*, 241–255. doi: 10.1207/s1532799xssr1003_3

Rinck, M., Gamez, E., Diaz, J. M., & de Vega, M. (2003). Processing of temporal information: Evidence from eye movements. *Memory & Cognition, 31*, 77–86. doi: 10.3758/BF03196084

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition, 35*, 213–235. doi: 10.1017/S0272263112000861

Shears, C., Miller, V., Ball, M., Hawkins, A., Griggs, J., & Varner, A. (2007). Cognitive demand differences in causal inferences: Characters' plans are more difficult to comprehend than physical causation. *Discourse Processes*, *43*, 255–278. doi: 10.1080/01638530701226238

Shimizu, H. (2012). *Generation of Japanese EFL readers' bridging inferences: The effects of text characteristics and L2 reading proficiency* (Unpublished doctoral dissertation). University of Tsukuba, Japan.

Singer, M. (1980). The role of case-filling inferences in the coherence of brief passages. *Discourse Processes*, *3*, 185–201. doi: 10.1080/01638538009544487

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, *22*, 437–448. doi:10.1016/S0022-5371(83)90282-7

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: an eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*, 251–272. doi: 10.1177/0267658310382068

Sonbul, S. & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*, 121–159. doi:10.1111/j.1467-9922.2012.00730.x

Spargo, E. (1989). *Timed readings book one* (3rd ed.). New York, NY: McGraw-Hill.

Stein, N. L., & Glenn, C. C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing: Advances in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.

Suh, S. Y., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory & Language*, *32*, 279–300. doi:10.1006/jmla.1993.1015

Suzuki, Y. (2000). Shougakusei wa bun wo rikai, kioku suru baai ni online de suiron wo hatarakasete iruka [Do elementary school children make on-line instrumental inferences in reading?]. *Japanese Journal of Educational Psychology*, *48*, 1–11.

Tapiero, I. (2007). *Situation models and levels of coherence.* Mahwah, NJ: Lawrence

Erlbaum Associates.

Trabasso, T., & Sperry, L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, *24*, 595–611. doi:10.1016/0749-596X(85)90048-8

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630. doi:10.1016/0749-596 X(85)90048-8

Trabasso, T., & Wiley, J. (2005). Goal plans of action and inferences during comprehension of narratives. *Discourse Processes*, *39*, 129–162. doi:10.1080/0163853X. 2005.9651677

Trabasso, T., & Wiley, J. (2009). What happens at reunions? Exploring causal connections and their role in reunion effects. *Discourse Processes*, *46*, 269–308. doi: 10.1080/01638530902959455

Ushiro, Y. (2010). *Flexibility of updating situation models: Schema modification processes of Japanese EFL readers*. Tokyo, Japan: Kairyudou.

van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, *27*, 1–22. doi: 10.1016/0749-596X(88)90045-9

van den Broek, P. (1994). Comprehension and memory of narrative texts: Inference and coherence. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539–588). New York, NY: Academic Press.

van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. *Child Development*, *67*, 3010–3028. doi:10.2307/1131764

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*, 1081–1087. doi:10.3758/BF03206376

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*, 1665–1690. doi: 10.1007/s11145-011-9337-4

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* San Diego, CA: Academic Press.

Virtue, S., van den Broek, P., & Linderholm, T. (2006). Hemispheric processing of inferences: The influence of textual constraint and working memory capacity. *Memory & Cognition*, *34*, 1341–1354. doi:10.3758/BF03193276

Webb, S. (2008). Receptive and productive vocabulary size of L2 learners. *Studies in Second Language Acquisition*, *30*, 79–95. doi:10.1017/S0272263108080042

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *40*, 361–387. doi:10.1080/02724988843000168

Yokokawa, H. (Ed.). (2006). *Nihonjin Eigo Gakushusha no Eitango Shinmitsudo: Moji hen* [English word familiarity for Japanese EFL learners: Letter volume]. Tokyo, Japan: Kuroshio Shuppan.

Yoshida, M. (2003). Working memory capacity and the use of inference in L2 reading. *JACET (The Japan Association of College English Teachers) Bulletin*, *36*, 1–17.

Yoshida, M. (2012). The interplay of processing task, text type, and proficiency in L2 reading. *Reading in a Foreign Language*, *24*, 1–29.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1196–1207. doi:10.1037/0278-7393.22.5.1196

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*, 292–297.

Zwaan, R. A., & Madden, C. J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 283–288. doi:10.1037/0278-7393.30.1.283

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 386–397. doi:10.1037/0278-7393.21.2.386

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language and memory. *Psychological Bulletin*, *123*, 162–185. doi:10.1037/0033-2909.123.2.162

Appendices

Appendix 1

Motivational Inference Passages, Target Probes, and Comprehension Questions Used in Experiments 1, 2, 4, 5, and 6

1. 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. [SC] / lost his grip. [WC])[26] (Experiments 1 and 2) *Suddenly, the ball dropped in front of the bat and fell in the catcher's mitt.*[27] (Experiments 4, 5, and 6)

Target word:	*hit*	(Experiments 1, 2 and 4)
Target sentence:	*He hits the ball.*	(Experiment 5)
Question:	*Was the first batter a new boy on the team?*	(Yes)

2. Linda sat down to read the newspaper in her living room. It was Sunday, so she could spend some extra time reading the paper. She picked up the paper and searched for the entertainment section. The room was darker than she liked, so Linda went over to the (blinds. [SC] / balcony [WC].) (Experiments 1 and 2) *She reached out her hand for the switch near the blinds and turned on the light.* (Experiments 4, 5, and 6)

Target word:	*open*	(Experiments 1, 2 and 4)
Target sentence:	*She opens the blinds.*	(Experiment 5)
Question:	*Was Linda reading the paper on Sunday?*	(Yes)

3. Tonight, John was having a party for his friends. He was always the life of the party. John was trying to think of a fun way to entertain his guests tonight. When an (upbeat disco [SC] / old-fashioned love [WC]) song came on the radio, he dramatically went to the center of the room. (Experiments 1 and 2) *Taking off his jacket, he looked around and asked his guests to play arm wrestling with him.* (Experiments 4, 5 and 6)

Target word:	*dance*	(Experiments 1, 2 and 4)
Target sentence:	*He dances in the room.*	(Experiment 5)
Question:	*Was John having a party for his friends?*	(Yes)

[26] SC = strong constraint; WC = weak constraint.
[27] A disconfirming sentence is written in italics.

4. The junior basketball star ran down the court. He felt in good condition that day. His teammates had a hard time keeping up with him. He stopped in front of the basket and looked at (it. [SC] / the referee. [WC]) (Experiments 1 and 2) *Then, the coach of the basketball team suddenly stopped the game and approached him slowly.* (Experiments 4, 5 and 6)

Target word:	*throw*	(Experiments 1, 2 and 4)
Target sentence:	*He throws the ball.*	(Experiment 5)
Question:	*Did the junior basketball* star *feel in bad condition that day?*	(No)

5. Robert was tired of making pizzas for a living and had a strong desire to do something more creative. After work, Robert went to the crowded city park. Once there, he looked around for pleasing view. He saw the beautiful sunset over the lake, and took out his (notebook and colored pencils. [SC] / memo pad. [WC]) (Experiments 1 and 2) *After closing his eyes for a while, he returned the notebook and pencils to his bag and left.* (Experiments 4, 5 and 6)

Target word:	*draw*	(Experiments 1, 2 and 4)
Target sentence:	*He draws the picture.*	(Experiment 5)
Question:	*Did Robert go to the city park?*	(Yes)

6. The rival gangs met outside the school yard. Both of the gangs had promised to be less violent. The neighbors watched as the two gangs shouted at one another. Finally, one member went over to the rival gang and put up his (fists. [SC] / hand. [WC]) (Experiments 1 and 2) *Then, the boss of the gang caught him by the arm and told him to calm down.* (Experiments 4, 5 and 6)

Target word:	*fight*	(Experiments 1, 2 and 4)
Target sentence:	*They fight with each other.*	(Experiment 5)
Question:	*Were the two gangs friendly?*	(No)

7. It was the end of a long week at work for Mary. She finished up her work for the day and told her boss that she was going home. She knew exactly what she was going to do when she got there. When Mary arrived home, she (got her pajamas on and turned off the lights. [SC] / sat down on the sofa and made herself comfortable. [WC]) (Experiments 1 and 2) *Then she noticed that her hair was dirty, so she got out of bed and went to the bathroom.* (Experiments 4 and 5)

Target word:	*sleep*	(Experiments 1, 2 and 4)
Target sentence:	*She sleeps on the bed.*	(Experiment 5)
Question:	*Did Mary finish her work for the day?*	(Yes)

8. Dave and Penny had a difficult relationship. They were in the middle of a heated discussion when Dave noticed a beautiful woman pass by. Dave’s eyes followed the attractive woman down the street. Penny turned to Dave and quickly raised her hand toward his (cheek. [SC] / arm. [WC]) (Experiments 1 and 2) *Then her cell phone rang loudly, and she went to answer it, still staring at him.* (Experiments 4 and 5)

Target word:	*hit*	(Experiments 1, 2 and 4)
Target sentence:	*She hits his cheek.*	(Experiment 5)
Question:	*Did Dave see the beautiful woman on the street?*	(Yes)

9. James had lived in the same place for 10 years. During that time, a lot of trash had piled up in his garage. It was full of sentimental photographs, books, and old clothes. James entered the garage and (grabbed a mop and a brush. [SC]/ a box full of photos. [WC]) (Experiments 1 and 2) *Putting them on the pile of trash, he left the garage and locked the door.* (Experiments 4 and 5)

Target word:	*clean*	(Experiments 1, 2 and 4)
Target sentence:	*He cleans the garage.*	(Experiment 5)
Question:	*Had James lived there only for a few days?*	(No)

10. Mrs. Smith, whose husband had been dead for 5 years, really needed money. Her savings had been spent, and she was now barely getting by. She loved a beautiful ruby necklace that her husband had given her. She sadly decided to go ask for help at a (jewelry shop. [SC]/ friend's house. [WC]) (Experiments 1 and 2) *After talking with the jeweler, it was decided that she would be hired to work in the shop.* (Experiments 4 and 5)

Target word:	*sell*	(Experiments 1, 2 and 4)
Target sentence:	*She sells her necklace.*	(Experiment 5)
Question:	*Was Mrs. Smith a rich woman?*	(No)

11. Brad was looking for a present for his wife's birthday. He wanted to find something special for her, but he couldn't afford to buy anything nice. In the accessories department, he saw an expensive ruby ring sitting in a display. Seeing no salespeople or customers around, he quietly made his way to the display and (opened his bag. [SC] / examined it more closely. [WC]) (Experiments 1 and 2) *Writing down the price of the ring on his memo pad, he promised to buy it someday.* (Experiment 4)

Target word:	*steal*	(Experiments 1, 2 and 4)
Question:	*Was Brad looking for a popular book?*	(No)

12. This was the final quarter of the Tigers' last football game of the season against the rival team. There were only thirty seconds left on the clock. The athletes' hearts raced with excitement as the clock started up. The Tigers were (close to [SC] / far from [WC]) their opponents' end zone, on the (5- [SC] / 50- [WC]) yard line, and they had the ball. (Experiments 1 and 2)*When approaching the end zone, one of the players of the rival team got back the ball from the Tigers.* (Experiment 4)

Target word:	*win*	(Experiments 1, 2 and 4)
Question:	*Was it the Tigers' last game of the season?*	(Yes)

13. The policeman saw the suspect trying to run away through the back door of the Bank of America. The policeman knew that he had to do something fast. The policeman pulled out his gun and shouted at the

running suspect. The policeman aimed his gun directly at the suspect, (but he still wouldn't stop. [SC] / and he immediately stopped. [WC]) (Experiments 1 and 2) *A police car suddenly appeared in front of him and he was caught by some police officers.* (Experiment 4)

Target word:	*shoot*	(Experiments 1, 2 and 4)
Question:	*Did the suspect run away through the back door of the police office?*	(No)

14. Steve and Susan were having a romantic picnic in the park. After they finished eating, Steve looked at his beloved Susan. He felt very close to her at this moment. He (got down on his knees and pulled out a diamond ring. [SC] / bent over and picked a rose for her. [WC]) (Experiment 1) *He asked her to throw it away because it was the ring his previous girlfriend had given to him.* (Experiment 4)

Target word:	*propose*	(Experiments 1 and 4)
Question:	*Were they watching a movie in the room?*	(No)

15. After the rugby match, Justin's friends made fun of him for not knowing the rules. He joined his friends and joked about beating them next time. Next, Justin went to grab a drink from the cooler. With a big grin, he (shook and opened a bottle of soda, aiming it at his friends. [SC] / opened a bottle of soda, looking for his friends. [WC]) (Experiment 1) *Justin gave it to one of his friends, who drank it all without taking a breath.* (Experiment 4)

Target word:	*spray*	(Experiments 1 and 4)
Question:	*Was Justin an excellent hockey player?*	(No)

16. Several people were outside that day. The air was colder than usual, but that didn't bother Bill. Bill got ready to join his friends. He was looking forward to getting some exercise (out on the ice. [SC] / outside today. [WC]) (Experiment 1) *Then he got a phone call from his mother and was told to come back home immediately.* (Experiment 4)

Target word:	*skate*	(Experiments 1 and 4)
Question:	*Was the air colder than usual?*	(Yes)

17. Sheila often got angry with her employees when they missed deadlines or behaved incompetently. The employees were meeting today to discuss a report they had been working on for 3 months. At the meeting, Sheila's secretary passed out the report and Sheila began to look through it. Sheila's face turned red with anger and she looked directly at her secretary when she realized that several pages were missing. (Experiment 2) *Then she found the secretary's eyes full of tears and turned her eyes away from the secretary without saying anything.* (Experiments 5 and 6)

Target word:	*shout*	(Experiment 2)
Target sentence:	*She shouts at the secretary*	(Experiment 5)
Question:	*Were the employees meeting today?*	(Yes)

18. A married woman and a young man were kissing passionately in her bedroom when they heard her husband come through the main door. She had forgotten he would be home earlier than usual tonight. The young man jumped up in panic when the husband called out to the woman and realized he couldn't leave without being seen. He noticed a closet near the bed and ran across the room towards it. *While doing so, he hurt his toe badly on the edge of the bed and lay on the floor motionless.* (Experiments 5 and 6)

Target sentence:	*He hides in the closet.*	(Experiment 5)
Question:	*Were a married woman and a young man kissing in the bed room?*	(Yes)

19. The girl was severely depressed that evening. She failed to pass the university entrance exam, and her life seemed so empty. She went to the roof of the tall building and watched the lights of the city. With a sense of despair, she looked down at the street beneath her feet and stepped toward the edge of the roof. *Just then, a man suddenly grabbed her arm and pulled her back to him to save her life.* (Experiments 5 and 6)

Target sentence:	*She jumps off the building.*	(Experiment 5)
Question:	*Did the girl pass the university entrance exam?*	(No)

20. Today, Lola finished her work earlier than usual. She had wanted to get the latest novel of her favorite writer. On the way home, she went into a bookstore and bought the novel. As soon as she reached home, she lay back on the sofa and opened the book. *She then received a phone call from her friend and left home without knowing the content of the book.* (Experiments 5 and 6)

Target sentence:	*She reads the book.*	(Experiment 5)
Question:	*Did Lola go to the book store after work?*	(Yes)

21. Harry went to his favorite restaurant with his girlfriend. The dishes there were excellent, and he enjoyed the dinner. After dinner, he asked her to wait outside the restaurant. Harry then asked the waiter to bring him the check, and he got out his wallet. *He did not know what to do when he realized that there was no money in the wallet.* (Experiments 5 and 6)

Target sentence:	*He pays the check.*	(Experiment 5)
Question:	*Did Harry hate the dishes in the restaurant?*	(No)

22. The university library was crowded with students. The first semester examination at the university was to begin three days later. A male student walked into the library and looked for an empty table. After he found it, he sat there and opened his notebook. *Soon after, he put his head on the notebook and slept without writing anything in the book.* (Experiments 5 and 6)

Target sentence:	*He studies in the library.*	(Experiment 5)
Question:	*Were there many students in the library?*	(Yes)

Appendix 2

Consequence Inference Passages, Target Probes, and Comprehension Questions Used in Experiments 1 and 3

1. Barbara was driving on the highway late at night struggling to stay awake. She smoked a cigarette to fight the boredom of the long drive. After she finished her cigarette, Barbara threw it out of the car window. The cigarette landed on a pile of (dry leaves at [SC] / wet leaves on [WC]) the side of the road. (Experiments 1 and 3)

   Target word: *fire*
   Question: *Was Barbara driving on the highway late at night?* (Yes)

2. David and his dog were enjoying a nice, long walking on the beach. He couldn't imagine a better way to spend his summer vacation. David decided to take his shoes off and ran into the water. With his next step, he didn't notice (a piece of broken glass under his foot. [SC] / a beautiful seashell nearby. [WC]) (Experiments 1 and 3)

   Target word: *cut*
   Question: *Was David walking on the beach with his dog?* (Yes)

3. The director and the cameraman were preparing for the next scene. They were new in Hollywood and had a lot to learn. The crew set up the cameras next to the building. The actress stood (by a window on the 14th floor [SC] / on the edge of the 2nd floor [WC]) and suddenly fell to the ground. (Experiments 1 and 3)

   Target word: *die*
   Question: *Had the director and cameraman worked for many years in Hollywood?* (No)

4. Michael and his colleague, Tanya, were having lunch at a restaurant. Tanya ordered the special of the day. The waiter quickly served her meal. Her eyes became round in surprise when she looked at her plate and saw (a giant cockroach. [SC] / some uncooked food. [WC]) (Experiments 1 and 3)

   Target word: *scream*
   Question: *Did Michael and Tanya go out to have their lunch?* (Yes)

5. The airplane was in flight to Europe. The passengers knew that they should soon be approaching their destination. They looked out the window and saw a mountain range a few feet away from them. The passengers knew they were (too close and cried out in fear. [SC] / close and gave a big sigh. [WC]) (Experiments 1 and 3)

Target word: *crash*
Question: *Were the passengers in the train?* (No)

6. Maria's new car had manual gears and she felt a bit anxious about driving it. When she got home from school, she parked the car and went inside the house. Then she realized that she had forgotten to put the emergency brake on. As she looked outside, she saw that her car (was parked on a sharp hill. [SC] / had been parked in the driveway. [WC]) (Experiments 1 and 3)

   Target word: *move*
   Question: *Did she go to school by bus?* (No)

7. The graduate student was working on his report in the library. He decided to take his work home since it was getting late. After a long evening, he gathered all of his papers and books together. As he lifted his books, his (arms [SC] / back [WC]) suddenly became very weak. (Experiments 1 and 3)

   Target word: *drop*
   Question: *Was the graduate student working hard on his part-time job?* (No)

8. The passengers had paid a great deal of money to take an historic ship across the ocean. The ship cruised through the ice-cold water at a high speed. The ship's captain became upset as the ship drifted off course. All of a sudden, the captain heard a terrible sound as the ship ran into (a large iceberg. [SC] / a small piece of ice. [WC]) (Experiments 1 and 3)

   Target word: *sink*
   Question: *Was the ship moving through the ice-cold water?* (Yes)

9. Daniel was using his toy dump truck next to the living room window. He liked to fill the truck with sand and then dump the sand out the window, down onto the street below. As Daniel was dumping sand, his truck fell out of the third-floor window. He watched as the weak, wooden truck fell toward the (cement driveway. [SC] / bushes. [WC]) (Experiments 1 and 3)

   Target word: *break*
   Question: *Was Daniel playing his video games?* (No)

10. Lily's home was a century old and needed many repairs. The roof, in particular, was in poor condition. She noticed one evening that there were many holes in the roof that needed to be fixed. Lily became worried when she learned of (the heavy rainfall. [SC] / approaching hurricane. [WC]) (Experiments 1 and 3)

    Target word: *drop*
    Question: *Did Lily live in her new home?* (No)

11. As Jimmy was coming home one day, he ran into some of the kids from the neighborhood. They asked him if he wanted to play with them. They taught him a fun game that involved throwing balls at a target to get points. He missed, though, and hit the front part of a new car with (a baseball. [SC] / a snowball. [WC]) (Experiments 1 and 3)

Target word:	*damage*	
Question:	*Did Jimmy play with the kids from neighborhood?*	(Yes)

12. Jennifer was getting ready for her big date, so she started a bath. She added her favorite aromatic oils to the water. Before she could turn off the water, she got a phone call from a childhood friend. After (30 [SC] / 4 [WC]) minutes on the telephone, Jennifer realized her bath was still running. (Experiments 1 and 3)

Target word:	*fill*	
Question:	*Did she get a phone call from her grandmother?*	(No)

13. Ralph was late for school so he pushed a piece of bread into the old toaster. After a few minutes, he could see that the bread had not come out of the toaster. Ralph didn't have anything else to eat and was determined to eat the toast. He used (a metal knife [SC] / his finger [WC]) to pick out the toast, forgetting that (the plug of the toaster was still connected. [SC] / it was still hot. [WC]) (Experiments 1 and 3)

Target word:	*shock*	
Question:	*Was Ralph in a hurry to go school?*	(Yes)

14. Lisa knew that her sister would love the chocolate cake that she was making for her birthday. After mixing the cake batter, she put it into a square-shaped baking pan. Then she turned on the oven, set the timer, and put the cake in the oven. Lisa didn't realize that she had (set the oven temperature too high. [SC] / set the timer off by a few minutes. [WC]) (Experiments 1 and 3)

Target word:	*burn*	
Question:	*Did Lisa's sister love anything spicy?*	(No)

15. Henry was a very careless man. He rarely watched where he was going. Today he was in a hurry to get home. As he was heading home, he stepped (on some ice. [SC] / in some mud. [WC]) (Experiments 1 and 3)

Target word:	*slip*	
Question:	*Was Henry in a hurry to go home?*	(Yes)

16. Jessica was hunting around in the kitchen. She found several jars that were very pretty. One of them looked especially interesting, so she removed the top. As Jessica put her nose into the jar, she found that it was full of (spicy pepper. [SC] / sweet cinnamon. [WC]) (Experiment 1)

Target word: *sneeze*
Question: *Were there several pretty jars in the kitchen?* (Yes)

17. Steven and his wife had been married for 10 years. Today, Steven was angry with his wife because she had left home without washing the dirty dishes in the kitchen sink. He tried to cool down, but felt his anger rising. No longer able to control himself, he threw a delicate glass vase against the wall. (Experiment 3)

Target word: *break*
Question: *Had Steven and his wife been married for many years?* (Yes)

Appendix 3

Neutral Passages, Target Probes, and Comprehension Questions Used in Experiments 1, 2, 4, and 5

1. 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. (Experiments 1, 2, 4, and 5)

Target word:	*hit*	(Experiments 1, 2, and 4)
Target sentence:	*He hits the ball.*	(Experiment 5)
Question:	*Did the young boys disappear the day before?*	(Yes)

2. The restaurant owner was nervous. He had just purchased the restaurant. To attract large crowds, he had advertised free desserts. He was pleased to see a huge line forming outside. (Experiments 1, 2, 4, and 5)

Target word:	*open*	(Experiments 1, 2, and 4)
Target sentence:	*She opens the blinds.*	(Experiment 5)
Question:	*Had he advertised free desserts?*	(Yes)

3. The rain fell gently on the earth. It was late summer and the ground needed some moisture. Eventually, small pools of water grew into huge ponds. After several days of rain, the ground was all covered with mud. (Experiments 1, 2, 4, and 5)

Target word:	*dance*	(Experiments 1, 2, and 4)
Target sentence:	*He dances in the room.*	(Experiment 5)
Question:	*Were there huge ponds of water on the ground?*	(Yes)

4. At sunset, the escaped prisoner ran into an abandoned old house. He knew he shouldn't stay there too long that night. He had served the first 3 years in a maximum security prison. He used his sweater as a pillow. (Experiments 1, 2, 4, and 5)

Target word:	*throw*	(Experiments 1, 2, and 4)
Target sentence:	*He throws the ball.*	(Experiment 5)
Question:	*Did the prisoner run into the old car?*	(No)

5. The high school band had practiced for months. Their big performance was later that night. Each year, the band had a big spring concert for the parents. Many said it was their best concert of the year. (Experiments 1, 2, 4, and 5)

Target word:	*draw*	(Experiments 1, 2, and 4)
Target sentence:	*He draws the picture.*	(Experiment 5)
Question:	*Did the band have a big concert for the parents?*	(Yes)

6. Tom and Marsha decided that they wanted their son to appreciate the mysteries of life. They decided to give him a pet as a birthday present. They put up a sign in town asking if anyone had a baby animal to give away. They got a call from a man who said that his dog was going to have babies any day. (Experiments 1, 2, 4, and 5)

Target word:	*fight*	(Experiments 1, 2, and 4)
Target sentence:	*They fight with each other.*	(Experiment 5)
Question:	*Did Tom and Marsha decide to give their son many toys?*	(No)

7. Tracy and Mary had been very close friends for many years. They were both sad when Mary had to move away. They managed to visit each other whenever they had free time. Their visits typically started off with talk about the old days. (Experiments 1, 2, 4, and 5)

Target word:	*sleep*	(Experiments 1, 2, and 4)
Target sentence:	*She sleeps on the bed.*	(Experiment 5)
Question:	*Did Mary have to move away?*	(Yes)

8. The dog pulled at his lead and barked. He looked up at his owner with sad brown eyes. Next, the dog lay on the ground. His owner would still not let him off the lead. (Experiments 1, 2, 4, and 5)

Target word:	*hit*	(Experiments 1, 2, and 4)
Target sentence:	*She hit his cheek.*	(Experiment 5)
Question:	*Were the eyes of the dog sad and brown?*	(Yes)

9. Although he was only 15 years old, James was special. He had an amazing ability. He could play golf like a player twice his age. His parents hoped he would grow up to be like Tiger Woods. (Experiments 1, 2, 4, and 5)

Target word:	*clean*	(Experiments 1, 2, and 4)
Target sentence:	*He cleans the garage.*	(Experiment 5)
Question:	*Did James have an amazing ability to play soccer?*	(No)

10. On special occasions, the family had big outdoor parties. They lived on a large and beautiful estate only a mile from the ocean. Parties typically started off with a small gathering on the beach. Tonight, they would end the night with some seafood dishes. (Experiments 1, 2, 4, and 5)

Target word:	*sell*	(Experiments 1, 2, and 4)
Target sentence:	*She sells her necklace.*	(Experiment 5)

Question:	*Did the family have big parties inside the house?*	(No)

11. Brad and Fred had been working in New York City for six months. Brad's parents were coming to stay with them this weekend. He was trying to think about what fun tourist spots they would visit. Brad and Fred decided to take them to the Museum of Art first. (Experiments 1, 2, 4, and 5)

Target word:	*steal*	(Experiments 1, 2 and 4)
Target sentence:	*He hides in the closet.*	(Experiment 5)
Question:	*Were Brad's parents coming to Tokyo?*	(No)

12. Marsha looked outside her window and she saw a beautiful sunrise. She quickly put on a sweater and went outside. Marsha lived near the beach and loved the mornings. She liked the thought that the town was quiet. (Experiments 1, 2, and 4)

Target word:	*win*	(Experiments 1, 2 and 4)
Question:	*Did Marsha see a beautiful sunrise?*	(Yes)

13. The teenagers were cruising around the streets of their town. One of the boys had a nice new sports car. The girls in town were impressed by the fast car. The boys often rode around town in their shiny new car. (Experiments 1, 2, 4, and 5)

Target word:	*shoot*	(Experiments 1, 2 and 4)
Target sentence:	*He studies in the library.*	(Experiment 5)
Question:	*Did one of the boys have a nice new bicycle?*	(No)

14. The festival was a confusion of the smells of many different people and beer. Susan heard the German band from the back of the crowd. She liked German music because her parents always played it when she was young. Susan looked at the crowd and thought how her parents would have enjoyed the music, too. (Experiments 1, 4, and 5)

Target word:	*propose*	(Experiments 1 and 4)
Target sentence:	*She jumps off the building.*	(Experiment 5)
Question:	*Did Susan like American music?*	(No)

15. Janet's coffee table had recently become unstable. She planned to pick up some tools to fix the loose leg. Somehow she could never find the time to stop at the store after work. She always wanted to get home and watch television. (Experiments 1, 4, and 5)

Target word:	*spray*	(Experiments 1 and 4)
Target sentence:	*She reads the book.*	(Experiment 5)
Question:	*Did she buy the tools she wanted at the store?*	(No)

16. Dan worked at the convenience store on the late shift. At about 2:00am, an unhealthy looking man came in. He walked around the store for about 20 minutes. Finally, he bought some gum and left the store. (Experiments 1 and 4)

Target word:	*skate*	(Experiments 1and 4)
Target sentence:	*She reads the book.*	(Experiment 5)
Question:	*Did an unhealthy looking man come into the store at midnight?*	(Yes)

17. The schoolyard was empty. All the students were already on their summer vacation. The groundskeeper was also on vacation. It was strange to see the schoolyard with no students around. (Experiments 2 and 5)

Target word:	*shout*	(Experiment 2)
Target sentence:	*She shouts at the secretary.*	(Experiment 5)
Question:	*Are there any students on the school yard?*	(No)

18. The therapist was extremely bored with his job. He was tired of hearing everyone else's problems. He had seen 10 clients that day, back to back. He started thinking about early retirement. (Experiment 5)

Target sentence:	*He pays the check.*	(Experiment 5)
Question:	*Was the therapist tired of his hob?*	(Yes)

Appendix 4

Neutral Passages, Target Probes, and Comprehension Questions Used in Experiments 1 and 3

1. The author worked day and night on her new book. She seldom rested, and if she did, it was only for an hour or two. She had worked this way for years. Her family worried that she would soon develop health problems. (Experiments 1 and 3)

Target word:	*fire*	(Experiments 1 and 3)
Question:	*Did she work hard writing her book?*	(Yes)

2. The atmosphere on the remote island was becoming tense. The survivors had a hard time getting along with each other. They didn't enjoy sharing the island. After only two days, the survivors were ready to leave. (Experiments 1 and 3)

Target word:	*cut*	(Experiments 1 and 3)
Question:	*Were they ready to leave the island?*	(Yes)

3. The first thing Rebecca and Jessica did the day after graduation was to look for summer jobs. They wanted to find fun jobs that had fairly flexible hours. They read all of the jobs advertisements but could not find anything. So they decided to meet some friends for drinks at their favorite bar. (Experiments 1 and 3)

Target word:	*die*	(Experiments 1 and 3)
Question:	*Did Rebecca and Jessica find a summer job?*	(No)

4. The state volleyball team was hot. They had won all of their matches that year. And, so far, there had been no injuries. It seemed the volleyball team was on the road to greatness. (Experiments 1 and 3)

Target word:	*scream*	(Experiments 1 and 3)
Question:	*Was the volleyball team strong?*	(Yes)

5. The chef hurried into the kitchen. The food critic had just arrived. The chef made sure that the critic was served their very best bottle of red wine. The critic examined everything about the restaurant before leaving. (Experiments 1 and 3)

Target word:	*crash*	(Experiments 1 and 3)
Question:	*Was the critic served a bottle of water?*	(No)

6. Maria ran her fingers through her dirty hair. The coffee was brewing and she gave a deep sigh. Next, she headed out to get the morning paper. Wrapping a bathrobe tightly around her, she went out of the

door. (Experiments 1 and 3)

Target word:	*move*	(Experiments 1 and 3)
Question:	*Did Maria go out to get the morning milk?*	(No)

7. Summer was definitely in the air. Parents were walking slowly outside with their new babies. The park was filled with kids playing on the swings. Soon, it would be hot enough to swim in the lake nearby. (Experiments 1 and 3)

Target word:	*drop*	(Experiments 1 and 3)
Question:	*Were there many dogs in the park?*	(No)

8. The young couple had very little money. They had just got married and their budget was tight. The first thing they wanted to buy was a large stereo. They asked for a loan from the woman's parents and purchased one. (Experiments 1 and 3)

Target word:	*sink*	(Experiments 1 and 3)
Question:	*Did the young couple want a big stereo?*	(Yes)

9. Marie had wanted to be a doctor for as long as she could remember. When she was a child, she was very impressed with her father's medical things. Once he let her try one of his instruments. She was fascinated by the sound of her own beating heart. (Experiments 1 and 3)

Target word:	*break*	(Experiments 1 and 3)
Question:	*Did she want to be a teacher when she was a child?*	(No)

10. It was Daniel's daughter's first day of elementary school. He came home from work early to hear all about her day. She was the youngest of his three children. She had been very excited to finally be going to school just like her two older brothers. (Experiments 1 and 3)

Target word:	*drop*	(Experiments 1 and 3)
Question:	*Did Daniel come home late at night?*	(No)

11. Jimmy was a taxi driver in Chicago. His first passenger of the day was an expensively dressed man. He asked Jimmy to take him directly to the airport. At the airport, the man gave Jimmy a huge tip without saying a word. (Experiments 1 and 3)

Target word:	*damage*	(Experiments 1 and 3)
Question:	*Did the passenger want to go to the airport?*	(Yes)

12. Rudy and Jennifer had just become proud parents of a baby boy. They decided to join a parenting group at the Young Men's Christian Association. Neither of them knew much about raising a child. They were both only 19 and hadn't planned on having a child so young. (Experiments 1 and 3)

Target word:	*fill*	(Experiments 1 and 3)
Question:	*Were Rudy and Jennifer so old?*	(No)

13. Other kids at school often made fun of the awkward girl in school. The girl had tried to make friends, but at times it seemed hopeless. She was terribly shy and her mother dressed her in old clothes. To make matters worse, the little girl often smelled bad. (Experiments 1 and 3)

Target word:	*burn*	(Experiments 1 and 3)
Question:	*Did the girl have many school friends?*	(No)

14. Karen and Ruth had been working in the emergency room. They had gotten their nursing degree two years ago. This morning they watched the police bring in three young children. They had been hurt on the jungle gym at their school. (Experiments 1 and 3)

Target word:	*shock*	(Experiments 1 and 3)
Question:	*Did the police bring the three children to the emergency room?*	(Yes)

15. The woman entered the hall. Her reputation preceded her wherever she went. She was currently dating a handsome man half her age. Whispers followed her as she moved across the room. (Experiments 1 and 3)

Target word:	*slip*	(Experiments 1 and 3)
Question:	*Was the woman dating a handsome man?*	(Yes)

16. Rose called to her brother when she saw something by the river. He waved to his sister, who was standing at the river’s edge. While standing completely still, they saw a crocodile. It slid off a very large rock into the water. (Experiment 1)

Target word:	*sneeze*	(Experiment 1)
Question:	*Was Rose standing at the river’s edge?*	(Yes)

17. The person directly next to Steven handed him an answer sheet. Soon, this hard chemistry class would be over. Steven nervously read over his notes one last time. He repeated the chemical sequences aloud to himself. (Experiment 3)

Target word:	*break*	(Experiment 3)
Question:	*Did he read over his chemistry class notes?*	(Yes)

Appendix 5

Control Passages and Comprehension Questions Used in Experiment 6

1. 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 without knowing it was a forkball.[28] Suddenly, the ball dropped in front of the bat and fell in the catcher's mitt.

   Question: *Was the first batter a new boy on the team?* (Yes)

2. Linda sat down to read the newspaper in her living room. It was Sunday, so she could spend some extra time reading the paper. She picked up the paper and searched for the entertainment section. The room was darker than she liked, so Linda went over to the switch on the wall. She reached out her hand for the switch near the blinds and turned on the light.

   Question: *Was she reading the paper on Sunday?* (Yes)

3. Tonight, John was having a party for his friends. He was always the life of the party. John was trying to think of a fun way to entertain his guests tonight. He decided to find out who was the strongest man among the guests at this party. Taking off his jacket, he looked around and asked his guests to play arm wrestling with him.

   Question: *Was John having a party for his friends?* (Yes)

4. The junior basketball star ran down the court. He felt in good condition that day. His teammates had a hard time keeping up with him. However, some teammates complained to him about his selfish performance. Then, the coach of the basketball team suddenly stopped the game and approached him slowly.

   Question: *Did the junior basketball star feel in bad condition that day?* (No)

5. Robert was tired of making pizzas for a living and had a strong desire to do something more creative. After work, Robert went to the crowded city park. Once there, he looked around at the pleasing view. Then he took out his notebook and colored pencils, but he couldn't find any inspiration for drawing. After closing his eyes for a while, he returned the notebook and pencils to his bag and left.

   Question: *Did Robert go to the city park?* (Yes)

[28] The underlined part is different from experimental (disconfirming) passages.

6. The rival gangs met outside the school yard. Both of the gangs had promised to be less violent. The neighbors watched as the two gangs shouted at one another. When one of the gang members finally raised his fists, someone shouted “Stop!” Then, the boss of the gang caught him by the arm and told him to calm down.

    Question: *Were the two gangs friendly?* (No)

7. Sheila often got angry with her employees when they missed deadlines or behaved incompetently. The employees were meeting today to discuss a report they had been working on for 3 months. At the meeting, Sheila’s secretary passed around the report, and Sheila began to look through it. Although she realized that several pages were missing, she tried to point it out to her secretary as gently as possible. Then she found the secretary’s eyes full of tears and turned her eyes away from the secretary without saying anything.

    Question: *Were the employees meeting today?* (Yes)

8. A married woman and a young man were kissing passionately in the woman’s bedroom when they heard her husband come through the front door. She had forgotten he would be home earlier than usual tonight. When the husband called out to the woman, the young man jumped up in panic, realizing he couldn’t leave without being seen. Although he couldn’t see clearly inside the room due to the darkness, he tried to hide under the bed. While doing so, he hurt his toe badly on the edge of the bed and lay on the floor motionless.

    Question: *Were a married woman and a young man kissing in the bed room?* (Yes)

9. The girl was severely depressed that evening. She had failed to pass the university entrance exam, and her life seemed very empty. She went to the roof of the tall building and watched the lights of the city. As she stepped toward the edge of the roof with a sense of despair, she heard someone cry out her name. Just then, a man suddenly grabbed her arm and pulled her back towards him to save her life.

    Question: *Did the girl pass the university entrance exam?* (No)

10. Today, Lola had finished her work earlier than usual. She had wanted to buy the latest novel of her favorite writer. On the way home, she went into a bookstore and bought the novel. However, on reaching home, she remembered that she was going to have dinner with her friend. She then received a phone call from her friend and left home without knowing what the book was about.

    Question: *Did Lola go to the book store after work?* (Yes)

11. Harry went to his favorite restaurant with his girlfriend. The food there was excellent, and he enjoyed the dinner. After dinner, he asked her to wait outside the restaurant. When Harry got out his wallet and looked inside it, he was very upset. He did not know what to do when he realized that there was no money in the wallet.

Question: *Did Harry hate the dishes in the restaurant?* (No)

12. The university library was crowded with students. The first semester examination at the university was to begin in three days. A male student walked into the library and looked for an empty table. He found one and opened his notebook there, but he couldn't concentrate at all. Soon afterwards, he put his head on the notebook and went to sleep without writing anything in it.

    Question: *Were there many students in the library?* (No)

Appendix 6

Filler and Practice Passages, Target Probes, and Comprehension Questions Used in Experiments 1 to 6

Filler Texts

1. Donald was normally a sensitive male. The only exception was when there was a football game on TV. His eyes became fixated and he became aggressive. No one could speak to Donald during a game.

Target word:	*aggressive*	(Experiments 1 and 4)	*tetch*	(Experiments 2 and 3)
Target sentence:	*He watches on coffee.*			(Experiment 5)
Question:	*Was Donald usually a sensitive man?*			(Yes)

2. Lee was an Internet junkie. When he wasn't at work, he was surfing the Internet. He did all of his shopping and communicating on the Internet. Lee was beginning to lose touch with reality. (Experiments 1, 2 and 3) Actually, it was not unusual for him to spend all day in his room, without talking to anyone. (Experiments 4 and 5)

Target word:	*lose*	(Experiments 1 and 4)	*gaulle*	(Experiment 2)	*tark*	(Experiment 3)
Target sentence:	*He rides the reality.*					(Experiment 5)
Question:	*Was he reading books when he wasn't at work?*					(No)

3. Camp Snoopy was alive with activity. There was a huge celebration. It was Snoopy's 60th birthday. Children came from all over the world to wish Snoopy a happy birthday.

Target word:	*wish*	(Experiments 1 and 4)	*smukes*	(Experiment 2)	*dal*	(Experiment 3)
Target sentence:	*They cry the present.*					(Experiment 5)
Question:	*Did many children love Snoopy?*					(Yes)

4. Maui loved to surf. His real name wasn't Maui, but that's the name he went by. He got the nickname by spending 4 months of the year in Hawaii. Maui spent those months riding the waves.

Target word:	*surf*	(Experiments 1 and 4)	*pask*	(Experiment 2)	*clak*	(Experiment 3)
Target sentence:	*He swims in the table.*					(Experiment 5)
Question:	*Did he get his nickname "Maui" in Hawaii?*					(Yes)

5. Karate was always an interest of William's. He was somewhat shy and karate gave him confidence. Karate helped him in many other ways, too. Last summer, he earned his black belt. (Experiments 1, 2 and 3) Because he was famous throughout his town for his great strength, many people talked to him. (Experiments 4 and 5)

Target word:	*shy*	(Experiments 1 and 4)	*brote*	(Experiment 2)	*hibbe*	(Experiment 3)
Target sentence:	*He cooks the belt.*					(Experiment 5)
Question:	*Was William a good kick-boxer?*					(No)

6. The mailman enjoyed his work. He had been on the same route for almost 7 years. He loved meeting the people every day on his route. One day he delivered a gigantic package to Mr. Fletcher. (Experiments 1, 2 and 3) Since Mr. Fletcher was old, the mailman carried the package to the living room of his house. (Experiments 4 and 5)

Target word:	*route*	(Experiments 1 and 4)	*chush*	(Experiment 2)	*healt*	(Experiment 3)
Target sentence:	*He grows the paper.*		(Experiment 5)			
Question:	*Did the mailman have trouble with people on his route?*					(No)

7. The cat was sad. He was used to being the only family pet. Now, his owners had brought home another pet. The cat was jealous of Molly, the golden retriever.

Target word:	*pet*	(Experiments 1 and 4)	*treg*	(Experiment 2)	*plest*	(Experiment 3)
Target sentence:	*He swallows the home.*					(Experiment 5)
Question:	*Was the cat jealous of another family pet?*					(Yes)

8. The designer had a new dress ready for the Spring line. It was a white, silk strapless dress that went to the knees. She hoped that her boss would like the new design. A moment later, she watched as her boss entered the room. (Experiments 1, 2 and 3) She became very nervous when the boss sat down and told her to start her presentation. (Experiments 4 and 5)

Target word:	*design*	(Experiments 1 and 4)	*rov*	(Experiment 2)	*healt*	(Experiment 3)
Target sentence:	*She put off the sea.*					(Experiment 5)
Question:	*Was the new dress white, silk strapless one?*					(Yes)

9. Samantha had finally finished her Ph.D. Now, she was anxious to find a job. She wanted to be a history professor. After hard work, Samantha ended up getting a great job at the University of New Mexico. (Experiments 1, 2 and 3) She has taught history there for more than 5 years, and now she is an assistant professor. (Experiments 4 and 5)

Target word:	*finish*	(Experiments 1 and 4)	*tark*	(Experiment 2)	*litch*	(Experiment 3)
Target sentence:	*She carries the school.*					(Experiment 5)
Question:	*Did Samantha want to be a math teacher in high school?*					(No)

10. The children's eyes became huge as they entered the candy store. Their parents didn't usually let them have any sweets. This was a special occasion. It was the youngest child's fourth birthday, so they were allowed a treat.

Target word:	*special*	(Experiments 1 and 4)	*pand*	(Experiment 2)	*crell*	(Experiment 3)
Target sentence:	*They travel to the arm.*					(Experiment 5)

Question: *Did the parents usually give their children many sweets?* (No)

11. The woman waited outside her apartment for a taxi. She was going to meet her sister while her car was being repaired. Her sister was very impatient, and didn't like to wait for people. She hoped her taxi would arrive soon. (Experiments 1, 2 and 3) Unfortunately, it took 30 minutes for her taxi to get to her apartment because of a terrible traffic jam. (Experiments 4 and 5)

    Target word: *wait* (Experiments 1 and 4) *phod* (Experiment 2) *domplet* (Experiment 3)
    Target sentence: *She drives the room.* (Experiment 5)
    Question: *Was the woman going to meet her boyfriend?* (No)

12. The two women greeted each other in the park. They were glad to catch up with each other. One woman had gotten married only 2 months ago. She told her friend all about her honeymoon in the French Riviera. (Experiments 1 and 2) After talking a while, they left the park to have lunch together at a restaurant. (Experiments 4 and 5)

    Target word: *marry* (Experiments 1 and 4) *baze* (Experiment 2)
    Target sentence: *They drink the hamburger.* (Experiment 5)
    Question: *Were they talking at a restaurant?* (No)

13. Shannon loved Italian designer dresses, purses, and shoes. She knew they were too expensive for her budget, but she had to have the latest styles. She traveled to Italy last fall just to go shopping. Shannon was thrilled with all of the new clothes she found there.

    Target word: *travel* (Experiment 1) *frain* (Experiment 2) *cripe* (Experiment 3)
    Question: *Did 13. Shannon visit French last fall just for shopping?* (No)

14. Christie was in line at the grocery store. She noticed the rack of magazines next to her. On the cover of one was her favorite actor, Brad Pitt. She read the article about him as she waited in line.

    Target word: *actor* (Experiment 1) *cripe* (Experiment 2)
    Question: *Was there a long line of people at the grocery store?* (Yes)

15. The boy just had heart surgery. He was scared, but very relieved when it was over. His family anxiously entered into his hospital room. He was happy to see that they brought him his favorite toy from home.

    Target word: *enter* (Experiment 1) *lase* (Experiment 2)
    Question: *Was the boy relieved when the surgery was over?* (Yes)

16. Willie was a talented songwriter. He wrote beautiful but strange lyrics. Willie had a hard time making a living, however. His songs were too abstract for most audiences.

Target word: *abstract* (Experiment 1) *poub* (Experiment 2)
Question: *Did Willie have a hard time earning much money by his songs?* (Yes)

17. The flea market was held every first Sunday of the month. Kim liked to shop there as often as she could. She decorated her house with items from there. Kim even managed to find unique gifts for her friends there.

Target word: *prack* (Experiment 2)
Question: *Was the flea market held every first Sunday of the month?* (Yes)

18. The mechanic fixed the starter on Ralph's car. It took him several hours to finish the job. He was experienced, but the shop was very busy that day. He had to finish working on 5 other cars, before he could finally fix Ralph's car.

Target word: *teave* (Experiment 2)
Question: *Did the mechanic fix Ralph's motorbike?* (No)

19. The baseball game was about to start. Mike put out the drinks and chips. He hoped the Dallas Cowboys would win. Mike was a big fan of the team. Actually, he dreamed that he would be a member of the team in the future. (used for the off-line phase in Experiment 5)

20. The farmer had a tough season. It hadn't rained for several weeks. He feared that he would not have a good harvest. He hoped that he would not have to give up his farm. He had been married and had 7 children, who were 4 sons and 3 daughters. (used for the off-line phase in Experiment 5)

21. Morton wanted to provide security for his children. He invested in stocks and bonds. He watched his investments carefully. Morton wanted to leave his children with a large sum of money. (used for the off-line phase in Experiment 5)

22. The basket, filled with onions and potatoes, was getting heavy. Helena switched the basket over to her other arm. She thought how silly it was to have bought the heavy things first. Now she had to carry this stuff through the rest of the Farmer's Market. (used for the off-line phase in Experiment 5)

Practice Texts

1. Marlene was a terrible typist. She was a professor, so it was important for her to have this skill. She bought a typing tutorial program for her computer. Marlene practiced typing for 3 hours each day.

Target word:	*click*	(Experiments 1 and 4)	*skump*	(Experiment 2)	*fitch*	(Experiment 3)
Target sentence:	*She clicks the man.*					(Experiment 5)
Question:	*Was Marlene good at typing?*					(No)

2. It was Christmas time and all of the children were excited. Santa Claus was rumored to be at the mall. Children begged their parents to see him. Many children wanted their pictures taken with Santa.

Target word:	*present*	(Experiments 1 and 4)	*plest*	(Experiment 2)	*drobe*	(Experiment 3)
Target sentence:	*They close the cake.*					(Experiment 5)
Question:	*Did children want to see Santa Claus?*					(Yes)

3. The baseball game was about to start. Mike put out the drinks and chips. He hoped the Dallas Cowboys would win. Mike was a big fan of the team. (Experiments 1, 2, and 3) Actually, he dreamed that he would be a member of the team in the future. (Experiment 4)

Target word:	*chip*	(Experiments 1 and 4)	*pitcher*	(Experiment 2)	*small*	(Experiment 3)
Question:	*Was he watching a baseball game?*					(Yes)

4. The farmer had a tough season. It hadn’t rained for several weeks. He feared that he would not have a good harvest. He hoped that he would not have to give up his farm. (Experiments 1, 2, and 3) He had been married and had 7 children, who were 4 sons and 3 daughters. (Experiment 4)

Target word:	*rain*	(Experiments 1 and 4)	*sun*	(Experiment 2)	*potato*	(Experiment 3)
Question:	*Did the farmer have a good harvest?*					(No)

5. Louis was always on the go. He was a social worker. His main duty was to help refugee families. Louis helped immigrant families get settled in their new country.

Target word:	*help*	(Experiments 1 and 4)				
Question:	*Was his work to take pictures of refugee families?*					(No)

6. The old man examined the people walking by him on a busy corner downtown. He no longer had any place to live and he didn’t have any money. He never imagined that he would ever be homeless. The man huddled to keep warm.

Target word:	*poor*	(Experiment 2)	*cold*	(Experiment 3)		
Target sentence:	*He finds the food.*					(Experiment 5)
Question:	*Did he have no place to live?*					(Yes)

7. The flight was crowded and hot, and the flight attendants were rude. The flight was going from Chicago to Paris. One passenger stood up and demanded a glass of water. Next, another angry passenger began to stand up in protest also.

Target word:	*drobe*	(Experiment 2)	*prack*	(Experiment 3)	
Target sentence:	*They talk to each other.*				(Experiment 5)
Question:	*Was the flight comfortable for passengers?*				(No)

8. Morton wanted to provide security for his children. He invested in stocks and bonds. He watched his investments carefully. Morton wanted to leave his children with a large sum of money.

Target word:	*urab*	(Experiment 2)	
Question:	*Did he want more money?*		(Yes)

9. The basket, filled with onions and potatoes, was getting heavy. Helena switched the basket over to her other arm. She thought how silly it was to have bought the heavy things first. Now she had to carry this stuff through the rest of the Farmer's Market.

Target word:	*carrot*	(Experiment 2)	
Question:	*Was her basket empty?*		(No)

Practice Target Sentences for the MJT in Experiment 5

1. *He writes the letter.*
2. *They visit the country.*
3. *She walks in the school.*
4. *They sing at the hall.*
5. *He starts from here.*
6. *She stands on the stage.*
7. *They stop by the store.*
8. *She plays with her brother.*
9. *They sit on the juice.*
10. *They hear the door.*
11. *They practice the rain.*
12. *She washes the season.*
13. *He opens the glass.*
14. *She leaves for the weather.*
15. *They run to the hour.*

Appendix 7

Target Sentences for the SRT in Experiments 5 and 6

Target sentences		
Explicit:	コーチは少年達の野球技術をテストすることにした	
Inference:	少年は球を打った	
Inconsistent:	少年はグローブを磨いた	(Experiments 5 and 6)
Explicit:	Linda は新聞のエンターテインメント欄を探した	
Inference:	彼女はブラインドを開けた	
Inconsistent:	彼女は新聞を届けた	(Experiments 5 and 6)
Explicit:	John は友達のためにパーティーを開いた	
Inference:	彼は部屋でダンスをした	
Inconsistent:	彼はパーティーでは目立たぬようにしていた	(Experiments 5 and 6)
Explicit:	バスケットボール選手はその日調子が良かった	
Inference:	彼はボールを投げた	
Inconsistent:	彼はベンチで休んでいた	(Experiments 5 and 6)
Explicit:	Robert は混雑している公園に行った	
Inference:	彼は絵を描いた	
Inconsistent:	彼はピザを作ることに情熱を傾けていた	(Experiments 5 and 6)
Explicit:	ギャング達は暴力的にならないと約束していた	
Inference:	彼らはケンカした	
Inconsistent:	彼らは学校の庭で談笑した	(Experiments 5 and 6)
Explicit:	Mary はその日の仕事を終えた	
Inference:	彼女はベッドで眠った	
Inconsistent:	彼女は上司と口論になった	(Experiment 5)
Explicit:	Dave は美しい女性が通っていることに気がついた	
Inference:	Penny は彼のほほをぶった	
Inconsistent:	Dave は Penny と仲良くした	(Experiment 5)
Explicit:	James は同じ場所に 10 年間住んでいた	
Inference:	彼はガレージを掃除した	
Inconsistent:	彼は古い服を売ることにした	(Experiment 5)
Explicit:	Mrs. Smith は 5 年前に夫を亡くした	
Inference:	彼女はネックレスを売った	
Inconsistent:	彼女は宝石店で指輪を盗んだ	(Experiment 5)

(continued)

Explicit:	社員たちは今日会議をしていた	
Inference:	彼女は秘書を怒鳴った	
Inconsistent:	彼女は資料を破り捨てた	(Experiments 5 and 6)
Explicit:	Susan はその夜ひどく落ち込んでいた	
Inference:	彼女はビルの屋上から飛び降りた	
Inconsistent:	彼女はきれいな夜景に感動した	(Experiments 5 and 6)
Explicit:	Harry は彼女とお気に入りのレストランへ行った	
Inference:	彼は代金を支払った	
Inconsistent:	彼は店員に文句をつけた	(Experiments 5 and 6)
Explicit:	図書館は学生たちで混雑していた	
Inference:	彼は図書館で勉強した	
Inconsistent:	彼はテーブルの上を拭いた	(Experiments 5 and 6)

Appendix 8

Profiles of Target Probe Words Used in Experiments 1 to 5

Target	Number of words	Frequency (JACET 8000 level)	Yokokawa's (2006) Familiarity	Experiment
hit	3	1	6.16	1, 2, 4, 5
open	4	1	6.77	1, 2, 4, 5
dance	5	1	6.13	1, 2, 4, 5
throw	5	1	5.56	1, 2, 4, 5
draw	4	1	5.19	1, 2, 4, 5
fight	5	1	6.29	1, 2, 4, 5
sleep	5	1	6.43	1, 2, 4, 5
clean	5	1	6.27	1, 2, 4, 5
sell	4	1	5.61	1, 2, 4, 5
steal	5	3	4.27	1, 2, 4
win	3	1	6.15	1, 2, 4
shoot	5	2	5.91	1, 2, 4
propose	7	3	5.33	1, 4
spray	5	1	6.48	1, 4
skate	5	8	not listed	1, 4
shout	5	1	5.29	2, 5
hide	4	2	5.14	5
jump	4	1	5.75.	5
read	4	1	6.58	5
pay	3	1	5.98	5
study	5	1	6.48	5
fire	6	2	5.18	1, 3
cut	3	1	6.23	1, 3
die	3	1	5.84	1, 3
scream	6	2	5.18	1, 3
crash	5	2	5.35	1, 3
move	4	1	6.27	1, 3
drop	4	1	5.27	1, 3
sink	4	2	3.96	1, 3
break	5	1	6.05	1, 3
drop	4	1	5.27	1, 3
damage	5	1	6.36	1, 3
fill	4	1	4.91	1, 3
shock	5	2	5.90	1, 3
burn	4	2	4.51	1, 3
slip	4	2	4.70	1, 3
sneeze	6	not listed	not listed	1

Note. In Experiment 5, these words were the main verbs of the MJT Targets.

Appendix 9

Profiles of Target Probe Sentences Used in Experiment 5

Target	Number of words	Number of syllables	Number of letters	Lexical frequency
He hits the ball.	4	4	13	1.00
She opens the blinds.	4	5	17	1.25
He dances in the room.	5	6	17	1.00
He throws the basketball.	5	6	21	2.40
He draws the picture.	4	5	17	1.00
They fight with each other.	5	5	18	1.00
She sleeps on the bed.	5	5	17	1.00
She hits his cheek.	4	4	14	1.25
He cleans the garage.	4	5	17	1.50
She sells her necklace.	4	5	19	2.40
She shouts at the secretary.	5	7	23	1.20
He hides in the closet.	5	6	18	2.60
She jumps off the building.	5	6	22	1.00
She reads the book.	4	4	15	1.00
He pays the check.	4	4	14	1.00
He studies in the library.	5	8	21	1.20

Note. Lexical frequency is the mean JACET 8000 level of words included in the target sentence.

Appendix 10

Profiles of Target Words Used for the Word-Recall Task in Experiment 2

Target	Number of letters	Frequency (JACET 8000 level)	Yokokawa's (2006) familiarity	Part of speech
legal	5	2	3.95	adjective
factor	6	2	4.65	noun
theory	8	1	5.61	noun
major	5	1	6.23	adjective
income	6	2	4.66	noun
structure	9	2	4.83	noun
occur	5	1	4.65	verb
method	6	1	4.55	noun
response	8	2	5.08	noun / verb
data	4	2	5.66	noun
assume	6	2	3.63	verb
benefit	7	2	4.71	noun
percent	7	1	6.07	noun
approach	8	1	5.67	noun / verb
identify	8	2	5.42	verb
role	4	1	5.00	noun
define	6	2	4.25	verb
concept	7	2	5.18	noun
specific	7	2	3.80	adjective
evidence	8	2	4.81	noun
indicate	8	2	4.43	verb
vary	4	2	4.51	verb
source	6	1	4.45	noun
similar	7	1	4.86	adjective
context	7	2	4.06	noun
estimate	8	2	3.77	verb
research	8	1	5.72	noun