Constructing Mental Representations of Textual Topic Structure Among Japanese EFL Readers

A Dissertation

Submitted to the University of Tsukuba

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Linguistics

Yoshinobu MORI

2017

Abstract of the Dissertation

Constructing Mental Representations of Textual Topic Structure Among Japanese EFL Readers

by

Yoshinobu MORI

Successful reading comprehension requires readers to construct discourse-level comprehension, beyond the understanding of individual words and sentences. To build globally coherent comprehension of texts, readers not only need to understand individual pieces of important text information, but they must also be able to establish links among this information. However, it has been found that English-as-a-foreign-language (EFL) readers have difficulty linking distant sentences in texts, which prevents them from achieving globally coherent comprehension (e.g., Morishima, 2013).

Theoretical frameworks of discourse comprehension commonly hypothesize that reading comprehension is a cognitive process of building coherent mental representations of texts (e.g., Gernsbacher, 1990; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; McKoon & Ratcliff, 1992). However, such a coherence-building process is difficult for EFL readers in the case of expository texts on unfamiliar topics. To build globally coherent comprehension of expository texts, readers are required to understand *topic structure*, which consists of individual topics (e.g., the major topic representing the whole text and subtopics representing individual paragraphs) and the hierarchical relations among them (e.g., Hyönä & Lorch, 2004).

Therefore, the present study investigated the construction of mental representations of

textual topic structure among Japanese EFL readers. Previous empirical studies have revealed how first-language (L1) readers linked the major topics and subtopics during reading and represented the topic structure in their text memory (e.g., Hyönä & Lorch, 2004; Hyönä, Lorch, & Kaakinen, 2002; Hyönä & Nurminen, 2006). In addition, previous research on second-language (L2) reading has demonstrated that L2 readers have difficulty linking topics beyond paragraphs (e.g., Ushiro, Nakagawa, Kai, Watanabe, & Shimizu, 2008). However, L2 and EFL research on this topic has been limited, and it is still unclear what the source of difficulty is and how to best support L2 reader understanding of topic structure.

Considering the importance and difficulty of globally coherent comprehension of texts in L2 and EFL reading, examining how Japanese EFL readers understand topic structure has both theoretical and pedagogical value. Therefore, the present study consists of two empirical studies that examined reader understanding of topic structure (Study 1) and the effect of educational interventions on promoting reader understanding of topic structure (Study 2).

Study 1 included a total of three experiments (i.e., Experiments 1–2B). Experiment 1 examined whether Japanese EFL readers represented topic structure in their text memory. The participants read several expository texts, each including major topics representing the whole texts and subtopics representing individual paragraphs. Both immediately following the reading session and two weeks after, the participants answered a written recall task where either the major topics, supporting details, or no information (control condition) was presented as recall cues. The result demonstrated that the recall rates for the subtopics were higher when the major topics were presented than when no information was presented, on both the immediate and delayed recall tasks. This suggests that the participants represented the topic structure in their text memory when the major topics were presented as retrieval cues.

Experiment 2A examined whether Japanese EFL readers understood topic structure while reading. After reading each expository text, the participants answered a primed recognition task. On the computer screen, one of the priming sentences representing the major topics, supporting details, or topically related but non-explicit information (control condition) was presented, and then the participants performed the recognition task with the target words representing the subtopics. However, the results were inconsistent between the correct response rates and times: While the correct response rates demonstrated reader understanding of topic structure, the correct response times did not.

To examine whether the inconsistent results were an outcome of a specific priming task or a more general phenomenon, Experiment 2B was conducted using a lexical decision task instead of the recognition task. The results indicated that neither the correct response rates nor correct response times indicated reader understanding of topic structure. Combined with the results of Experiment 2A, the correct response times did not indicate understanding of topic structure, regardless of the priming task. This suggests that the difficulty of understanding topic structure during EFL reading is a general tendency. On the other hand, because the correct response rates indicated understanding of topic structure in Experiment 2A alone, the participants might have understood it by referring back to their text memory and then reconstructing it in the recognition task.

Study 2 included two experiments (Experiments 3–4). Experiment 3 examined the effect of reading instructions on reader understanding of topic structure. The participants read expository texts sentence by sentence at their own pace and then answered the written recall task. To assess reader understanding of topic structure during reading and in the post-reading task, the explicitness of the major topics was manipulated. In addition, to explore the effect of the reading instructions, the participants were instructed to read the texts in order to answer comprehension questions (i.e., standard condition) or write outlines (i.e., instruction condition) after reading. The reading times and recall rates of the subtopics indicated that the participants failed to understand topic structure during reading or in the post-reading task, regardless of the given instructions.

Experiment 4 examined the effect of task engagement on reader understanding of topic structure. The participants engaged in a think-aloud process while reading the expository texts that included the major topics and subtopics and then answered the written recall task. They read the texts to answer comprehension questions (i.e., standard condition) or they wrote an outline at the time of reading (i.e., task condition). The think-aloud comments indicated that writing outlines helped the higher-proficiency group to selectively reread the major topics and subtopics of the texts and identify the hierarchical relations between them. This group tended to complete the outline task after understanding the entire texts, and it was found that they understood topic structure while completing the outline task. In addition, the high-proficiency readers' outline and recall performances indicated that this group better understood topic structure and represented it in their text memory. However, writing outlines was not an effective method of promoting understanding of topic structure for the lower-proficiency group.

The main findings of the present study can be summarized into the following three points: (a) Japanese EFL readers have difficulty understanding topic structure while reading and representing it in their text memory without any aids (e.g., retrieval cues or engaging in a productive task); (b) although just giving the outline instructions was not effective, writing an outline helped the higher-proficiency readers to understand topic structure after reading and represent it in their text memory; and (c) the lower-proficiency readers had difficulty sufficiently adapting their cognitive processes to educational interventions to understand topic structure.

The present dissertation provides deep theoretical insights into whether and how Japanese EFL readers link major topics and subtopics beyond paragraphs in expository reading. Furthermore, the present findings also possess pedagogical implications for supporting globally coherent comprehension among EFL readers using effective and appropriate educational interventions.

Acknowledgements

I would like to thank many people who supported me over the course of my research at the Graduate School of Humanities and Social Sciences at the University of Tsukuba. Although I would like to express my appreciation to everyone who helped me write this dissertation, I would like to acknowledge the following people.

First, my deepest gratitude goes to my academic supervisor, Professor Yuji Ushiro at the University of Tsukuba. He prompted my interest in English education and reading in a second language. I received many insightful comments from him and his seminar, and I learned the importance of bridging theory and practice. These comments always enlightened me greatly and helped me make progress in my research. I also appreciate his encouragement when I had difficulty writing this dissertation.

I am indebted to Professor Hirosada Iwasaki at the University of Tsukuba. He gave me suggestions regarding the role of vocabulary in reading and I learned academic writing in his lectures. Professor Akira Kubota at the University of Tsukuba also advised me from the perspective of task-based language learning. His comments helped me to consider cognitive processes of engaging in educational interventions. Furthermore, I would like to express my heartfelt appreciation to my external supervisor, Professor Masashi Negishi at the Tokyo University of Foreign Studies. He gave me suggestions in terms of English learners' proficiency and pedagogical implications, which assisted me in revising the earlier version of this dissertation.

Many thanks also go to my senior associates in Professor Ushiro's seminar, particularly Dr. Shingo Nahatame, Dr. Yusuke Hasegawa, Dr. Yukino Kimura, Dr. Akira Hamada, and Dr. Natsumi Tanaka. They advised me from various viewpoints, and I found their vast insight to be valuable to my research. Moreover, I am grateful to all my junior associates for their help in conducting my research and writing this dissertation. I would also like to thank my colleague, Mr. Masaya Hosoda, who worked with me during most of my research and shared both good and hard times with me.

I am also obliged to the participants in my experiments for this dissertation. I could not finish it without their cooperation.

Finally, a special thanks goes to my family, who always stood by me and offered their support.

Yoshinobu MORI

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1
Introduction

1.1 Background of the Present Study

One of the primary goals of reading is to understand the overall message that the writer is intending to convey in the text. To achieve this goal, readers first need to comprehend pieces of important text information, such as the topics of paragraphs, sections, and entire texts. However, even if readers can understand individual topics, it does not mean that they sufficiently understand the text as a whole. Furthermore, the readers are required to construct well-structured mental representations of *topic structure*, which consists of the individual topics and hierarchical links among them (e.g., Hyönä & Lorch, 2004). For example, when the major topic (e.g., environmental problems) is explored through the subtopics of the supporting paragraphs (e.g., global warming, air pollution), global warming and air pollution should be understood as different forms of environmental problems. While the understanding of topic structure contributes to globally coherent text comprehension, the failure to do so leaves the readers with limited or fragmented mental representations of the text.

To date, previous empirical research has demonstrated that first-language (L1) readers linked topics during reading and then represented the topic structure in their text memory (e.g., Hyönä & Lorch, 2004; Hyönä, Lorch, & Kaakinen, 2002; Hyönä & Nurminen, 2006). On the other hand, despite the significance of globally coherent text comprehension, it has been pointed out that second language (L2) and English as a foreign language (EFL) readers have difficulty linking text information beyond the paragraph level (e.g., Johns & Mayes, 1990; Ushiro, Nakagawa, Kai, Watanabe, & Shimizu, 2008). However, because there are few previous studies examining EFL reader understanding of topic structure, the mechanisms of globally coherent text comprehension in EFL reading and the possible causes of its difficulty have not been sufficiently clarified. Therefore, the present study investigates the cognitive processes of EFL readers that are relevant to the understanding of topic structure.

Both the importance and the difficulty of understanding text topic structure indicate that EFL readers need scaffolding. However, in current EFL educational settings, reading instructions mainly target local-level comprehension (e.g., the anaphoric or semantic relations within a sentence or between close sentences). On the other hand, educational interventions aimed at supporting globally coherent text comprehension among EFL readers are quite limited.

Previous studies on L1 reading have explored the effects of educational interventions (e.g., reading instruction, engagement in tasks) on reading comprehension, and found that L1 readers flexibly adapt their cognitive processes to the interventions (e.g., van den Broek, Lorch, Linderholm, & Gustafson, 2001; Redford, Thiede, Wiley, & Griffin, 2012). However, the past literature on educational interventions to support globally coherent text comprehension in EFL reading is limited (e.g., Kimura, 2014, 2015a; Ushiro et al., 2017). Other studies have reported that EFL readers sometimes fail to adapt their reading processes to educational interventions (e.g., Horiba, 2013; Yoshida, 2012). Hence, the present study also explored the effect of educational interventions on globally coherent text comprehension in EFL reading, focusing on reader understanding of topic structure.

The present study will provide theoretical insights into the mechanisms and challenges of EFL reader construction of globally coherent links among topics. Further, the pedagogical implications of the present study's findings will be discussed, with suggestions for educators regarding how to reduce the difficulty of globally coherent text comprehension in EFL reading, and how to better facilitate the understanding of topic structure.

1.2 Organization of This Dissertation

This dissertation consists of five chapters: the Introduction (Chapter 1), Literature Review (Chapter 2), Study 1 (Chapter 3), Study 2 (Chapter 4), and General Discussion and Conclusion (Chapter 5).

Chapter 2 reviews the previous research on reading comprehension. Specifically, theories and models of text comprehension (e.g., mental representations of texts), previous research on locally and globally coherent text comprehension among L1 and L2 readers, the effects of educational interventions (i.e., giving reading instructions, engaging in tasks) on reading comprehension, and methodologies to assess reader understanding of topic structure are reviewed. This chapter concludes with a summary of the findings and limitations of the previous studies.

To clarify the construction of mental representations of textual topic structure among Japanese EFL readers, this dissertation conducted three experiments in Study 1 and two experiments in Study 2. Figure 1.1 illustrates the overview of the five experiments in the two studies. Study 1 investigated reader understanding of topic structure during reading and in a post-reading task. Study 2 explored the effects of educational interventions on reader understanding of topic structure.

Chapter 3 reports on Study 1, which examined topic structure processing during EFL reading and reader memory of it. Experiment 1 investigated whether the Japanese EFL readers represented the topic structure in their text memory. The Japanese university students read a set of expository texts, each including the major topics representing the whole texts and subtopics representing the paragraphs. They answered the written recall task immediately after reading and two weeks later, and either the major topics, supporting details, or no information were provided as recall cues. The recall rates of the subtopics were compared among the conditions regarding the recall cues to assess the readers' memory of the hierarchical links between the major topics and subtopics.

Study 1: Understand Topic Structure		**Factors**	**Measurements**
Experiment 1	Memory of topic structure	L2 reading proficiency Recall cues Recall time	Cued recall task
Experiment 2A	Topic structure processing during reading	L2 reading proficiency Priming sentences	Recognition task
Experiment 2B			Lexical decision task
Study 2: Effects of Educational Interventions		**Factors**	**Measurements**
Experiment 3	Effect of giving outline instructions	L2 reading proficiency Major-topic explicitness With/without the instructions	Self-paced reading Cued recall task
Experiment 4	Effect of engagement in outline task	L2 reading proficiency Reading conditions With/without task engagement	Thinking aloud Outline task Cued recall task

Figure 1.1. Overview of the five experiments of the present study.

Experiments 2A and 2B investigated whether the Japanese EFL readers understood topic structure during reading. The Japanese university students read expository texts and answered a priming task. After reading each text, one of the priming sentences representing the major topic, a supporting detail, or topically related but non-explicit information was presented, followed by the target words representing the subtopics. The participants engaged in the recognition task and lexical decision task with the target words in Experiments 2A and 2B, respectively. The correct response times and rates were compared according to the type of priming sentence, which evaluated their during-reading comprehension of the topic structure.

Chapter 4 reports on Study 2, which explored the effects of educational interventions on reader understanding of topic structure. Experiment 3 investigated whether the outline instructions helped the Japanese EFL readers understand topic structure during reading and in a post-reading task. The Japanese university students engaged in sentence-by-sentence, self-paced reading of the expository texts, in which the explicitness of the major topics was manipulated. The participants were instructed to read the texts to answer comprehension questions (i.e., standard condition) or write outlines (i.e., instruction condition). After reading, the participants answered the written recall task. The reading times and recall rates of the subtopics were compared as a function of the major-topic explicitness (i.e., explicit vs. non-explicit) and reading condition (i.e., standard vs. instruction) to analyze the instruction effect on reader understanding of topic structure during reading and in the post-reading task.

Experiment 4 examined whether writing outlines helped the Japanese EFL readers understand the topic structure during reading and in a post-reading task. The Japanese university students engaged in a think-aloud process while reading the expository texts, focusing on the major topics and subtopics, and then performed the written recall task. The participants read to answer comprehension questions (i.e., standard condition), or they wrote an outline at the time of reading (i.e., task condition). The proportions of the think-aloud comments and the recall quality were compared between the reading conditions (i.e., standard vs. task). The quality of the outlines themselves were also analyzed.

Chapter 5 discusses the general results of the five experiments and presents the study's conclusions regarding the construction of mental representations of textual topic structure among Japanese EFL readers. Next, the limitations of the present study are discussed and suggestions for future research are made. Finally, the pedagogical implications of the study's findings are suggested.

Chapter 2
Literature Review

2.1 Reading Comprehension

2.1.1 Theories and Models of Reading Comprehension

By reading a text, a reader constructs a meaningful representation of the text information in their mind (e.g., Johnson-Laird, 1983; Kintsch, 1998). This is referred to as a *mental representation*, and can be defined as “some change in the way the mind views the world as a result of reading a text,” or “some sort of trace of the text read” (Kintsch, 2004, p. 1271). Because a mental representation is constructed through the interaction between a reader and a text, both the reader’s prior knowledge and the content of the text are integrated into a mental representation (e.g., van Dijk & Kintsch, 1983). Theoretical frameworks of reading comprehension commonly assume that reading is an activity that constructs a coherent mental representation. This section will overview theoretical frameworks of building coherent mental representations in text comprehension: the construction-integration (CI) model (e.g., Kintsch, 1988), the structure building framework (e.g., Gernsbacher, 1990), and the landscape model (e.g., van den Broek, Risden, Fletcher, & Thurlow, 1996).

The construction-integration model

The CI model was first proposed by Kintsch (1988), but it was developed based on his preceding framework of multilevel mental representations (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). This framework hypothesizes that three levels of mental representations are constructed as a result of reading texts: *surface memory*, propositional *textbase*, and *situation models*. As the first level of reading comprehension, surface memory is a representation of the same words and order among them as the original text. Surface memory is the shallowest level of mental representations, as it rapidly fades from the reader's memory. At the second level, the propositional textbase represents the meaning extracted from the network of propositions, each consisting of one or more arguments and a predicate (i.e., the relationships among the arguments; Kintsch & van Dijk, 1978, p. 367), and *argument overlap* integrates the propositions into a coherent network. At the third and deepest level, the situation model is the most enduring representation of the overall meaning of the text. To construct a situation model, readers need to actively generate inferences in reading, and link text information with their prior knowledge.

In addition, Zwaan and his colleagues further classify situation models into three types: the *current model*, *integrated model*, and *complete model* (Zwaan, 1999; Zwaan & Radvansky, 1998). Whereas the current model is a representation of the focal sentence, the integrated model integrates all representations constructed up until the immediately preceding sentence. The integration of the current model into the integrated model is called *updating*. After reading all the sentences, integrating all the representations constructed from the text results in the complete model, which is encoded into the readers' long-term memory.

Based on the perspective of multilevel mental representations, Kintsch (1988, 1998, 2004) proposed the CI model to explain how readers integrate text information with their knowledge to construct situation models. This model regards text comprehension as a cycle of two processes: *construction* and *integration*. During the first phase of construction, text information and readers' relevant knowledge are activated automatically in a bottom-up manner, which creates a network of propositions. Because this activation is not regulated by the conscious and strategic control of the readers, irrelevant or inconsistent association is activated as well. During the second phase of integration, the activation spreads across the network until it settles. Specifically, a concept remains activated if it is connected to many other concepts or it is related to the focal sentence, reading goal, or reading perspective. If a concept does not satisfy these conditions (i.e., it is irrelevant or inconsistent with information activated during the construction phase), it is deactivated and disappears from the readers' mental representations.

The structure building framework

The structure building framework (Gernsbacher, 1990, 1997) states that "the goal of comprehension is to build coherent mental representations or *structures*" (Gernsbacher, 1997, p. 265). This framework hypothesizes three component processes of building coherent structures: *foundation laying*, *mapping*, and *shifting*. First, readers need to lay a foundation for their mental structures. This foundation laying often occurs at the beginning of a unit of a text (e.g., a paragraph), and the mental effort for it increases the reading times for the initial sentence of that unit. Once readers lay the foundation, they develop their mental structures by mapping the incoming text information onto it. When the incoming information coheres with or relates to the previous information, a related concept is activated in their mental structures (i.e., *enhancement*). However, when the incoming information is less coherent with or less related to the previous information, the related concept is suppressed (i.e., *suppression*). In this case, a different concept is activated, directing readers to shifting and laying a new foundation. The structure shifting results in longer reading times for sentences that initiate new topics than for sentences that do not create such a topic shift. Britton (1994) further elaborated the structure building framework. Britton assumed that readers basically expand on a topic of a unit of text. However, when they realize that the unit is complete, readers wrap up the information written in it into the *gist* (i.e., *unitizing*) and suppress the activation of information other than the gist to ensure the cognitive resources necessary to process the next unit.

The structure building framework differs from the CI model concerning the processes that occur when readers encounter irrelevant information. The structure building framework assumes that readers actively suppress such irrelevant information in their mental structure. On the other hand, the CI model hypothesizes that readers automatically deactivate the irrelevant information, which is not affected by conscious or strategic processes. However, although there are some differences between these theoretical frameworks, both explain how readers build coherence between pieces of text information.

The landscape model

van den Broek and his colleagues proposed the landscape model (e.g., van den Broek, Risden, et al., 1996). Their model simulates the activation of concepts that occur in reading processes, based on the assumption that reading is a cycle of processing sentences or propositions. Concepts are activated in readers' minds based on the following four sources: (a) the current text, (b) carry-over from the previous cycle (i.e., concepts readily available in readers' working memory), (c) reinstatement from prior cycles, and (d) activation of readers' background knowledge. These variables determine the activation patterns of concepts, which form their situation models and text memory.

The landscape model hypothesizes two mechanisms of concept activation: automatic *cohort activation*, and strategic *coherence-based retrieval*. Cohort activation is an automatic and memory-based process. When a new concept is activated in the reader's working memory, it sends signals to all the concepts in their long-term memory (i.e., text memory, prior knowledge). Relevant concepts resonate with the signals and are reactivated in working memory. As a result, relevant concepts activated together are linked and form the cohort. On the other hand, coherence-based retrieval is a strategic process to retrieve readers' long-term memory or physically look back to prior text to achieve *standards of coherence*, or "the types and strength of coherence a reader aims to achieve in a particular reading situation" (e.g., van den Broek, Beker, & Oudega, 2015, p. 98). When readers perceive that automatically activated concepts are not sufficient to meet their standards of coherence, they engage in strategic processes (i.e., inference generation). Readers' standards of coherence are not always stable and are affected by reader factors (e.g., working memory capacity, reading goals), text factors (i.e., text genres, difficulty), and tasks. The landscape model hypothesizes that the strategic process depends on readers' standards of coherence, although it is similar to the CI model in that it assumes that readers use automatic and strategic processes.

2.1.2 Theories of Coherence-Building Processes

To capture the overall meaning of a text, readers need to understand not only the relations between close sentences, but also those between distant sentences. Otherwise, their comprehension is limited to a narrow part of the text and fragmented. Such relations between close and distant sentences are called *local coherence* and *global coherence*. Although their definitions differ among researchers, Graesser, Singer, and Trabasso (1994) stated as follows:

> Local coherence refers to structures and processes that organize elements, constituents, and referents of adjacent classes or short sequences of clauses. Global coherence is established when local chunks of information are organized and interrelated into higher order chunks. (p. 371)

To build coherent comprehension, readers need to link pieces of text information using *inference*. Although there are various definitions of inference appearing in different studies, van den Broek, Risden, and Husebye-Hartmann (1995) defined it as follows: "Inferences refer not only to the addition of facts or events as nodes in the mental representation of the text, but also the identification of relations between nodes that already exist" (p. 353). The next section will review the theoretical assumptions concerning coherence-building inference according to the *constructionist theory* (Graesser et al., 1994) and *minimalist hypothesis* (McKoon & Ratclif, 1992), and will provide an overview of their empirical findings as well.

The constructionist theory

Graesser et al. (1994) proposed the constructionist theory, which predicts which type of inference readers generate during reading. The principal tenet of their theory is the effortful, active, and strategic process called the *search after meaning*, which is based on the *reader goal assumption*, *coherence assumption*, and *explanation assumption*. In the reader goal assumption, readers construct meaning representations in accordance with their goals. The representations and goals are set at the deep level, such as situation models. In the coherence assumption, readers attempt to achieve both locally and globally coherent representations. In the explanation assumption, readers attempt to explain the reasons for text information.

The constructionist theory classifies inference into 13 types (see Table 2.1). This theory assumes that inferences that build local coherence (i.e., *local inference*; Cases 1 to 3) and global coherence (i.e., *global inference*; Cases 3 to 6) are generated during reading (i.e., *on-line inference*) because these inferences are necessary for establishing a meaningful representation. On the other hand, *elaborative inferences* (i.e., Cases 7 to 11), which lead to more detailed understandings of texts but are not always critical for achieving a coherent representation, are not assumed to occur during reading but after reading is completed (i.e., *off-line inference*). However, there are exceptions for on-line inference; that is, on-line inference is not generated during reading (a) when readers detect deficits in a text, such as a lack of global coherence or message; (b) when readers do not have the necessary background knowledge to build explanatory and global coherence; or (c) when reading goals do not require readers to construct meaningful situation models (e.g., reading for proofreading).

Table 2.1

Types of Inference (Adopted From Graesser et al., 1994, p. 375)

Case	Brief description
1. Referential	A word or phrase is referentially tied to a previous element or constituent in the text (explicit or inferred)
2. Case structure role assignment	An explicit noun phrase is assigned to a particular case structure role, e.g., agent, recipient, object, location, time.
3. Causal antecedent	The inference in on a causal chain (bridge) between the current explicit action event, or state and the previous passage context.
4. Superordinate goal	The inference is a goal that motivates an agent's intentional action.
5. Thematic	This is a main point or moral of the text.
6. Character emotional reaction	The inference is an emotion experienced by a character, caused by or in response to an event or action.
7. Causal consequence	The inference is on a forecasted causal chain, including physical events and new pants of agents. These inferences do not include the character emotions in class 6.
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.
9. Instrument	The inference is an object, part of the body, or resource used when an agent executes an intentional action.
10. Subordinate goal-action	The inference is an ongoing state, from the times frame of the text, that is not 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.
11. State	The inference is an ongoing state, from the time frame of the text, that is not 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.
12. Emotion of reader	The inference is the emotion that the reader experiences when reading a text.
13. Author's intent	The inference is the author's attitude or motive in writing.

Minimalist hypothesis

In opposition to the constructionist theory, McKoon and Ratcliff (1992) proposed the minimalist hypothesis. While the constructionist theory emphasizes the effortful, active, and strategic processes involved in text comprehension, the minimalist hypothesis assumes that text comprehension is an effortless, passive, and automatic process. Because of this difference, assumptions regarding on-line inference generation are different in the two theoretical frameworks (see Table 2.2). According to the minimalist hypothesis, inferences are normally generated during reading and are based on readily available information in the readers' working memory alone: well-known information from readers' general knowledge and information described explicitly in the text being read. Hence, the minimalist hypothesis assumes that local inferences (i.e., Cases 1 to 3) are normally generated during reading because information that is "no farther apart in the text than one or two sentences" (McKoon & Ratcliff, 1992, p. 441) is still readily available in readers' working memory. On the other hand, global inferences (i.e., Case 4 to 6) are not automatically generated in normal reading situations because they require reactivating more distant preceding information in the readers' long-term memory. Global inferences are generated during reading when the text lacks local coherence or when readers engage in strategic reading to achieve specific reading goals.

Thus, the constructionist theory and minimalist hypothesis have different assumptions concerning on-line inference at local and global levels. Although both frameworks state that generating local inferences is an on-line process, generating global inferences is an on-line process according to the constructionist theory but an off-line process according to the minimalist hypothesis.

Table 2.2

Predictions of On-Line Inferences From the Minimalist Hypothesis and Constructionist Theory

Case	Function	Minimalist	Constructionist
1. Referential	Local	X	X
2. Case structure role assignment	Local	X	X
3. Causal antecedent	Local	X	X
4. Superordinate goal	Global		X
5. Thematic	Global		X
6. Character emotional reaction	Global		X
7. Causal consequence	Elaboration		
8. Instantiation of noun category	Elaboration		
9. Instrument	Elaboration		
10. Subordinate goal-action	Elaboration		
11. State	Elaboration		
12. Emotion of reader			
13. Author's intent			

Note. This table was adapted from Graesser et al. (1994, p. 384). X = on-line prediction.

To date, many empirical studies have examined global inference in L1 narrative reading, using the *inconsistency-detection paradigm* (e.g., Albrecht & O'Brien, 1993). This paradigm assumes that if readers attempt to establish coherence during reading, they should experience comprehension difficulty when encountering input inconsistent with the prior context (i.e., *inconsistency effect*). For example, in Albrecht and O'Brien (1993), L1 university students read a target sentence (e.g., "Mary ordered a cheeseburger and fries") seven sentences after reading a consistent or inconsistent sentences (e.g., "Mary enjoyed eating anything that was quick and easy to fix" vs. "Mary, a health nut, has been a strict vegetarian for 10 years"). The sentence-by-sentence, self-paced reading times for the target sentences were longer after reading the inconsistent contexts than the consistent ones, suggesting that the participants attempted to establish globally coherent comprehension during narrative reading. Moreover, the recall rates of the contexts and target sentences (i.e., text information relevant to the inconsistencies) were higher for the inconsistent texts than the consistent texts, indicating that global coherence was represented in the participants' text memory. These results were replicated in later L1 research (Hakala & O'Brien, 1995).

The above findings were further expanded. In eye-tracking studies, L1 university students showed more frequent and longer look-backs in inconsistent texts than in consistent ones (Poynor & Morris, 2003; Rinck, Gámez, Díaz, & de Vega, 2003). This suggests that the participants reactivated distant preceding information by physically looking back to the prior text. Moreover, several studies have reported the effect of reading proficiency on globally coherent comprehension among L1 elementary school students (van der Schoot, Reijntjes, & van Lieshout, 2012) and L1 university students (Long & Chong, 2001). In these studies, good readers established both local and global coherence during reading, but poor readers were only able to maintain local coherence.

In the field of L2 reading research, studies adopting the inconsistency detection paradigm have been quite limited compared to L1 research. Morishima (2013) compared coherence building by L1 readers and EFL readers. The result showed that both groups achieved coherence between adjacent sentences. However, when a single sentence was inserted between the sentences, EFL readers were not able to maintain coherence although L1 readers were able to do so. A similar result was also observed in Ushiro, Nahatame, et al. (2016, Experiment 2). In their study, Japanese university students linked two sentences when a single sentence was inserted, but they failed to do so when four sentences were inserted. These L2 studies reasoned that their EFL participants had difficulty allocating sufficient cognitive resources to achieve globally coherent comprehension during reading because their insufficient lower-level processes (e.g., word recognition, syntactic parsing) consumed most of their available cognitive resources. However, Ushiro, Mori, et al. (2016) revealed that Japanese university students achieved both local and global coherence in their text memory. Combined with the difficulty of EFL reading, they suggested that their participants established coherent comprehension by reorganizing or reconstructing their text memory in a recall task. That is, more cognitive resources were likely available for globally coherent comprehension after the resource-demanding lower-level processes of literal text comprehension had been completed. This idea is supported by the findings of other L2 studies (Hosoda, 2014; Nahatame, 2013), in which Japanese EFL readers failed to generate inferences (i.e., causal antecedent and consequence) during reading but succeeded in doing so in post-reading tasks.

In sum, the abovementioned studies found that L1 readers built globally coherent comprehension during reading and represented it in their text memory, although their reading proficiency level might have affected it. In L2 reading, globally coherent comprehension was difficult during reading probably because of insufficient cognitive resources. It is possible that more cognitive resources were available for globally coherent comprehension in post-reading tasks than during reading. However, L2 research has not fully explored this research topic. Moreover, because narrative texts have been used as materials in most of the past studies using the inconsistency-detection paradigm, it is not clear whether the past findings can be applied to expository reading. Therefore, the next section will review the features of expository reading at the global level.

2.1.3 Features of Expository Reading at the Global Level

There are two major text genres: narrative and expository. Graesser, Li, and Feng (2015) stated that "Narrative text tells a story with characters, events, places, and things that are familiar to the reader." (p. 301). In contrast, according to Coté, Goldman, and Saul (1998), expository texts "frequently present concepts and relations that readers do not already know" (p. 6). Although there are other differences between narrative and expository texts, one of the most critical differences that might affect coherence-building processes at a global level is text structure. Hence, this section will overview the features of text structure in expository texts, in comparison with narrative texts.

Regarding the structure of narrative texts, the story grammar aims to describe the typical order of text information in narrative texts (Thorndyke, 1977). For example, Thorndyke (1977) analyzed the story grammar of a simple narrative, proposing that it consists of the following four elements: the *setting*, *theme*, *plot*, and *resolution*. The setting describes stative information of the time, location, and main characters. The theme states or implies the goal the main characters attempt to achieve in the subsequent plot. The plot consists of many episodes, each containing the *subgoal*, *attempt*, and *outcome*. The subgoal is the method necessary for achieving the goal (i.e., theme), followed by the attempts that are actions for achieving the subgoal. The outcome is the consequence of the attempts, describing whether or not the actions or attempts satisfy the subgoal. Finally, the resolution concludes the story in terms of its theme.

A large body of empirical studies have explored the assumptions of the story grammar. For example, in Carrell (1984b), it was found that L2 university students better understood the narrative texts of the standard version of a narrative than the interleaved version. Whereas the standard version first described all the elements of Episode 1 and then those of Episode 2, each element was interleaved across Episodes 1 and 2 (i.e., from the beginnings of Episodes 1 and 2 to the outcomes of Episodes 1 and 2). Moreover, past research has indicated that readers recall more higher-level information than lower-level information in the hierarchical story structure (Horiba, van den Broek, & Fletcher, 1993; van den Broek, Lorch, & Thurlow, 1996). These findings from previous studies demonstrate that readers construct mental representations of narrative texts in accordance with the story grammar.

On the other hand, the text structure of expository texts is more complicated than that of narrative texts (Meyer, 1975; Meyer & Freedle, 1984). Meyer and Freedle (1984) classified the text structure of expository texts into *collection*, *description*, *causation*, *problem-solution*, and *comparison*. They defined collection as an associative list of concepts and description as a specific type of associative list of concepts regarding a superordinate idea (i.e., topic). When the collection of many concepts describes a superordinate topic, such a text structure is called the *collection of descriptions* (e.g., Carrell, 1984a). The causation texts represent groups of concepts that are related not only chronologically but also causally. The problem/solution texts first present problems about a certain topic and then solutions to the problems. Finally, the comparison texts are organized based on opposing viewpoints about a specific topic.

Empirical studies have examined the effects of text structure on expository reading. For example, Carrell (1984a) revealed that L2 readers recalled less information from expository texts representing a collection of descriptions than other text structures. She reasoned that her participants had difficulty understanding collections of descriptions because they are less organized than other structures, such as the structures of causation, problem/solution, and comparison. That difficulty has been replicated by later studies, although these studies slightly differed in their understandings of other text structures (Bohn-Gettler & Kendeou, 2014; Gaith & Harkouss, 2003). In addition, Carrell (1984a) found that the participants recalled more text information using the same structure as the texts; however, this was observed among a limited number of participants. Hence, although it is difficult to detect the text structure of expository texts, it has been found that reader awareness of text structure positively affects text memory (Carrell, 1984a, 1992; Martinez, 2002).

As reviewed above, expository texts have more varying and complicated text structures than narrative texts. Because of such differences, text comprehension, including making global inferences, might differ according to the text genre. Hence, past empirical research has compared narrative and expository reading. For example, Horiba (2000, Experiment 1) showed that L2 readers recalled fewer main ideas from expository texts than from narrative texts. This result was replicated by Yoshida (2012), who also proved that the above tendency was observed in a delayed recall task as well. Further, regarding on-line reading processes, Shimizu (2015) compared Japanese university students' coherence-building processes at a global level during narrative and expository reading. The think-aloud protocols suggested that her participants allocated fewer cognitive resources to generating global inferences during expository reading than narrative reading, regardless of their L2 reading proficiency. The abovementioned findings suggest that L2 readers have more difficulty building global coherence during expository reading and display poorer text memory than during narrative reading. Therefore, the next section will review globally coherent comprehension of expository texts to better understand what makes it so difficult for L2 readers.

2.2 Understanding Topic Structure in Expository Texts

2.2.1 Frameworks of Understanding Topic Structure

As reviewed in 2.1.2, previous studies have examined globally coherent comprehension mainly in narrative reading (e.g., Albrecht & O'Brien, 1993). On the other hand, limited studies have investigated this topic in expository reading (e.g., Shimizu, 2015). Thus, this section first overviews the frameworks that can explain coherence-building processes at the global level in expository reading, and then reviews related findings from previous empirical studies.

In the case of expository texts, building globally coherent comprehension requires readers to represent each topic in their mental representations and the relation of one topic to one another within a text (referred to as topic structure; e.g., Hyönä & Lorch, 2004). Lemarié, Lorch, Eyrolle, and Virbel (2008) defined *topic* as “a concept or theme that is the focus of elaboration in a section of text” (p. 33), although this definition differs slightly among researchers. There are various levels of topics that represent a paragraph, a chapter, and the whole text. Topics at higher and lower levels are often referred to as major topics and subtopics, respectively; in other words, *major topics* and *subtopics* are relative terms. For example, Hyönä (1994) used experimental texts that consisted of an introduction of a main theme (i.e., major topic), detailed coverage, and a conclusion, and described the relation between the major topic and subtopics as follows:

> Each of the nine target paragraphs in the detailed coverage unit specified one subtopic that was a particular instantiation of the main theme. For example, in the Athens text, one subtopic dealt with the status of the slaves, another with leisure time activities, a third with how laws were passed. (p. 78)

Lorch and Lorch (1985), who examined understandings of relations between the major topics representing entire texts and the subtopics representing paragraphs, hypothesized that good readers represent topic structure during reading and retrieve it after reading in a top-down manner. First, when readers identify the main topic during reading, they can use it as a context to interpret incoming information. Second, when the readers encounter a new topic during reading, they retrieve their evolving mental representations of the topic structure, and then integrate the new topic into the appropriate location of that hierarchical representation. Finally, their mental representations of topic structure support text memory retrieval in post-reading tasks. Specifically, retrieving a topic provides access to information organized under that topic. The first process is similar to foundation laying and mapping in Gernsbacher's structure building framework, and the second process is similar to shifting.

In contrast, van Dijk and Kintsch's model (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983) assumes that readers understand topic structure in a bottom-up manner. In their model, readers integrate individual supporting details written in a text (called *micropropositions*) into a superordinate proposition called the *macroproposition*, representing a larger part of the text. Further, readers continue to integrate macropropositions into a superordinate macroproposition until a single macroproposition is constructed to represent the entire text. This recursive process constructs macropropositions at different hierarchical levels (e.g., subtopics of a paragraph, and the major topic of the entire text), each is hierarchically linked to other macropropositions to build globally coherent mental representations throughout the text (i.e., referred to as *macrostructure*). Figure 2.1 illustrates the hierarchy of macrostructure.

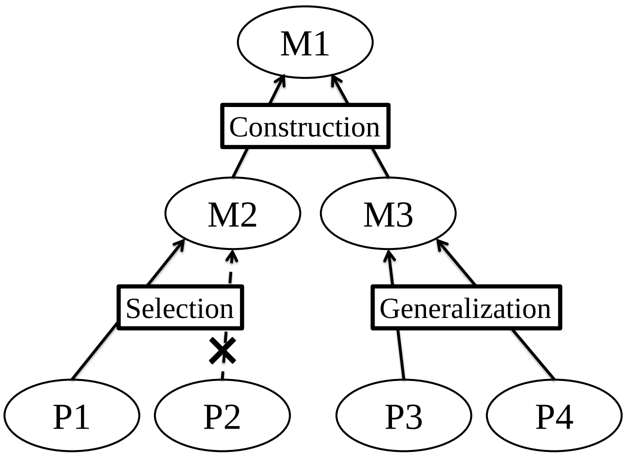

Figure 2.1. The hierarchical macroprostructure and macrorules. The letters P and M represent micropropositions and macropropositions, respectively.

Many macropropositions are explicitly written in texts as titles and topic sentences. However, macropropositions are not always explicitly written in texts. To achieve globally coherent comprehension in such cases, readers need to infer the implicit macroproposition from the subordinate propositions on their own. To understand explicit and implicit macropropositions, readers need to use three summarization rules (referred to as*macrorules*): *selection*, *generalization*, and *construction*. By adopting the selection rule, readers distinguish explicit macropropositions from detailed text information. Next, the generalization rule integrates propositions including subordinate words into a macroproposition including a superordinate word (e.g., *pencil*, *eraser*, and *scissors* → *stationary*). Finally, the construction rule summarizes the subordinate propositions into a superordinate proposition (e.g., *write the address*, *add a stamp*, and *drop the letter into a mailbox* → *send the letter*). Whereas the selection rule contributes to understanding explicit macropropositions, the generalization and construction rules lead to inferring implicit macropropositions.

In sum, both frameworks proposed by Lorch and Lorch (1985) and van Dijk and Kintsch (1983) hypothesize that readers establish globally coherent representations of topic structure throughout the text. However, these frameworks differ on the following points. First, whereas Lorch and Lorch (1985) assumed a top-down processing style like the structure building framework (Gernsbacher, 1990), van Dijk and Kintsch (1983) hypothesized a bottom-up processing style. Moreover, the former describes reading processes when individual topics are explicit in the text, but the latter explains reading processes when topics are implicit as well. These frameworks are likely to explain reader understanding of topic structure.

2.2.2 Empirical Findings on Understanding Topic Structure

Building globally coherent comprehension of expository texts requires readers to link the propositions in texts with one another (e.g., Hyönä & Lorch, 2004). To understand the hierarchical relations among propositions, readers need to discriminate and organize propositions at different hierarchical levels such as supporting details, subtopics representing a paragraph, and major topics representing a larger part of the text. Additionally, they must extract implicit topics from subordinate propositions in texts (van Dijk & Kintsch, 1983). In light of these demands on readers, this section will review past empirical studies as a function of hierarchical levels and the explicitness of topics.

Both L1 and L2 studies have found that readers understand the hierarchical relations among topics and their supporting details within a paragraph. For example, Guindon and Kintsch (1984) conducted a recognition task immediately after L1 university students read expository paragraphs with and without topic sentences, which were target probes in the recognition task. The correct response times and rates demonstrated that the explicit and implicit topics were activated in the participants' minds, suggesting that they understood the topics during reading. A similar result was also observed in Mori (2015), who adapted Guindon and Kintsch's recognition task for Japanese university students.

Previous studies using the reading time method more directly examined building global coherence between paragraph topics and supporting details. In Lorch (1993), L1 university students were interrupted in the middle of reading a paragraph of an expository text. Before restarting reading, the topic sentence that initiated the paragraph was or was not presented. The reading times for the first sentences (i.e., supporting details) after resumption were shorter when the topic sentences were presented, showing that the participants were linking the supporting details with the explicit topics during reading. Additionally, Ritchey (2011) and Kimura (2013) demonstrated that L1 and EFL undergraduates read the last sentences of paragraphs faster when these summarized the paragraphs than when they did not. Their findings suggested that L1 and EFL readers generalized the supporting details into implicit paragraph topics during reading.

Additionally, performance on post-reading tasks can also be indicative of reader understanding of the hierarchical links within a paragraph. Goldman, Saul, and Coté (1995) indicated that L1 university students produced more topic sentences of paragraphs than supporting details in reproduction tasks (i.e., recall and summary tasks) after reading. Similar results have also been obtained in L2 research (e.g., Kim, 2001; Ushiro et al., 2008). Ushiro et al. (2008) revealed that Japanese university students included more paragraph topics than supporting details in their summaries, regardless of whether they were explicitly or implicitly suggested in the text.

On the other hand, L1 and L2 research has often shown that it is difficult to integrate subordinate propositions into an implicit topic representing a larger part of the text beyond the paragraph level. For example, Brown and Day (1983) found that few readers were able to infer implicit topics beyond paragraphs, even in L1 reading. In L2 reading, think-aloud study (Kimura, 2015b) have shown that Japanese university students had difficulty inferring the overall themes of expository texts during reading. Moreover, previous research has also found that L2 university students were not able to produce implicit topics across paragraphs in summary tasks (Johns & Mayes, 1990; Ushiro et al., 2008).

As reviewed so far, past studies have suggested that L1 and L2 readers understand the hierarchical relations linking explicit and implicit topics with supporting details within paragraphs during reading and in post-reading tasks. On the other hand, it has been also shown that readers have difficulty integrating subordinate propositions and inferring implicit topics beyond paragraphs, regardless of being L1 or L2 readers. Considering these findings, readers might first need to aim at grasping the hierarchical relations between explicit topics across paragraphs such as the links between the subtopics of a paragraph and the major topics beyond it. This issue has been explored mainly in L1 reading research, and few studies have been conducted in L2 research. The following section will review empirical studies that investigated hierarchical linking among explicit topics.

Topics beyond paragraphs are explicitly described in varying forms such as overviews, headings, titles, and so on (Lemarié et al., 2008). For example, in Lorch, Lemarié, and Grant (2011b), L1 university students selected the major topics more often than the subtopics when required to identify the major topics (i.e., Experiment 3). Additionally, they located more shifts between major topics than subtopics when required to identify major topic shifts (i.e., Experiment 4). A series of experiments demonstrated that L1 readers understood the hierarchical relations between the major topics and subtopics.

Lorch, Lorch, Ritchey, McGovern, and Coleman (2001) also observed successful understanding in productive tasks. In their research, L1 university students summarized expository texts including topics at different hierarchical levels with or without various types of headings (e.g., differences in capitalization, italicization, or indentation). The participants produced text information from more paragraphs with headings than without headings. In the place of the summary task and headings, Lorch, Chen, and Lemarié (2012, Experiment 1) replicated this finding adopting the outline task and preview sentences that introduced topics in the subsequent sections of texts. In Lorch, Lemarié, and Grant (2011a, Experiments 1 and 2), presenting headings helped L1 readers to write appropriate outlines representing hierarchical topic structure.

Murray and McGlone (1997) investigated the during-reading processes of linking paragraph subtopics with major topics representing the entire texts. The researchers manipulated the introductory paragraphs, presenting general information alone or *topic overviews*, which initiate expository texts and explicitly identify what will be discussed in the following text, such as in the following example: “This passage will examine the aspects of Morinthia that account for its uniqueness. We will consider aspects such as Morinthia's geography, climate, major exports, major imports, inhabitants, and political system" (Murray & McGlone, 1997, p. 260). The reading times for topic sentences in the following paragraphs were shorter with the topic structures than without them, demonstrating that L1 readers were linking the paragraph subtopics with the major topics described in the topic overview.

Hyönä and Lorch (2004) expanded on Murray and McGlone's (1997) findings by recording eye movements during expository reading with or without headings representing two paragraphs. The L1 university students read the topic sentences of the paragraphs more smoothly with headings than without them, and they often looked back to preceding sentences from the beginning or end of paragraphs (i.e., boundaries between paragraphs). This suggests that the participants were linking the subtopics with the major topics (e.g., Gernsbacher, 1990; Lorch & Lorch, 1985) and wrapping up subordinate propositions after reading paragraphs (e.g., Britton, 1994; van Dijk & Kintsch, 1983), as the theoretical frameworks assume.

Further, a study by Cauchard, Eyrolle, Cellier, and Hyönä (2010) marked paragraphs representing topics at different hierarchical levels, using different types of headings (i.e., font size, font type, and degree of indentation). L1 university students read the texts in the normal condition or window condition, which allowed them to see a few current lines including one heading alone (i.e., they did not have access to the hierarchical information conveyed by the combination of different heading types). The participants looked back to the headings for shorter times and less frequently in the window condition than in the normal condition. The window interrupted reader understanding of hierarchical topic structure, suggesting that L1 readers normally understand the relations between the major topics and subtopics.

In addition to the evidence of during-reading comprehension, other past studies have demonstrated the hierarchically-structured nature of text memory. In Lorch and Lorch (1996), L1 university students read an expository text with or without textual signals (i.e., headings, an overview, or a summary). They presented cues in the recall task to remind the participants of the major topics of the entire text. The participants recalled text information from more paragraphs of the signaled text, compared to the non-signaled text. This result indicates that the L1 readers represented topic structure in their text memory. As Lorch and Lorch (1985) hypothesized, their mental representations of topic structure might have guided their recall; retrieval of the major topic provided access to information about the subtopics linked under the major topic. Similar findings were also found in other studies as well (Chambliss, 1995; Lorch & Lorch, 1985; Ritchey, Schuster, Allen, 2008; Surber & Schroeder, 2007).

The abovementioned studies indicated that L1 university students comprehended topic structure during reading and in post-reading tasks. Moreover, Chambliss and Murphy (2002) also suggested that less proficient readers were able to understand hierarchical topic structure. L1 fourth and fifth graders read an expository text and then completed a written recall task. The recall protocols from the fifth graders represented the hierarchical structure of the writer's argument throughout the text and the supporting data in the paragraphs, although the recall protocols from the fourth graders were less structured. This suggests that non-proficient readers had some ability to understand topic structure and then represented it in their text memory.

The above studies indicated that reader understanding of topic structure was better achieved through the completion of post-reading tasks. However, it should be noted that these studies did not consider which information in the sections of texts (i.e., paragraph subtopics or supporting details) was linked to the signaled major topics (e.g., Chambliss, 1995; Chambliss & Murphy, 2002; Lorch & Lorch, 1985; Lorch et al., 2001; Lorch et al., 2011a; Lorch et al., 2012; Ritchey et al., 2008; Surber & Schroeder, 2007). Specifically, in the scoring of post-reading tasks, it was determined that a topic was produced when a single proposition was produced, whether it was a paragraph subtopic or a supporting detail. To better understand the organization of readers' mental representations, it seems necessary to specify how propositions are linked in readers' mental representations.

Unlike L1 reading research, few L2 studies have explored reader understanding of the hierarchical relations between the major topics and text information in paragraphs. Although the focus was different from the present study, Ushiro et al. (2009) investigated whether presenting different titles affected summary writing by EFL readers. Their expository text described how the United Nations International Children's Emergency Fund (UNICEF) provided girls in developing countries with opportunities for education, and Soda Fall's (a woman's name) dream came true thanks to the aid. In the summary task, either "UNICEF Project" or "Soda Fall" was used as the text title. The participants' summaries included more propositions that were relevant to the given title and thus rated as important. This suggests that EFL readers are able to link the major topics (i.e., titles) with text information in the paragraphs relevant to the title. However, more studies are needed to clarify how L2 readers understand topic structure in expository reading.

2.2.3 Factors Affecting Globally Coherent Comprehension

Previous studies have suggested possible factors that might affect globally coherent comprehension, which can be divided into the categories of reader factors and measurement factors. Regarding reader factors, L1 studies have demonstrated the general tendency that readers' skills of building global coherence increase with their age (e.g., Brown & Day, 1983; Chambliss & Murphy, 2002; Kintsch, 1990; Williams, Taylor, & de Cani, 1984; Williams, Taylor, & Ganger, 1981; Winograd, 1981). Specifically, while children rely on eliminating unimportant pieces of information and selecting important information, mature readers can link and condense propositions. In particular, organizing information across paragraphs is a later developed skill, even in L1 reading.

In addition to the age of readers, previous research has reported the effect of reading proficiency on globally coherent comprehension. In L1 reading, van der Schoot, Vasbinder, Horsley, and van Lieshout's (2008) eye-tracking study revealed that fifth and sixth graders fixated on more important information longer than less important information during expository reading, which was correlated with discourse comprehension. In a post-reading task, Winograd (1984) found that poor child readers depended on simple selection and deletion of individual propositions in the summary task, whereas good child readers were more engaged in linking propositions.

In L2 reading, effect of L2 proficiency on globally coherent comprehension was inconsistent among past studies. For example, Kimura (2014) showed that Japanese university students with higher L2 reading proficiency better understood the themes of expository texts than lower-proficiency students. On the other hand, the effect of L2 reading proficiency did not appear in the summarization skills in Ushiro et al. (2009), suggesting that L2 reading proficiency was likely to affect literal comprehension more than building globally coherent comprehension. Furthermore, Johns and Mayes (1990) suggested the possibility that L2 proficiency partially influenced summarization strategies. Specifically, while the lower-proficiency group depended on more direct copies of the texts, the higher-proficiency group was more engaged in linking a few sentences. However, the participants failed to produce implicit main ideas, especially those across paragraphs, regardless of their L2 proficiency. Their study might indicate that L2 proficiency effect on globally coherent comprehension differs based on types of cognitive processes. Considering this view and the fact that Ushiro et al. examined summarization skills within paragraphs, it was also possible that L2 reading proficiency had a larger effect on coherent comprehension beyond paragraphs than within paragraphs. Because building globally coherent comprehension requires organizing larger parts of texts, the effect of L2 reading proficiency might appear more noticeably here.

Furthermore, the eye-tracking studies of Hyönä and his colleagues clarified that there are individual differences in reading behaviors (Hyönä et al., 2002; Hyönä, & Nurmine, 2006). Based on eye movements, L1 university students were categorized as follows: liner readers, who looked back less frequently; nonselective reviewers, who frequently looked back regardless of the information's relevance to the hierarchy of topic structure; and topic structure processors, who fixated and looked back at sentences longer and more frequently in accordance with the topic structure (e.g., look-backs to headings). Moreover, they found that the topic structure processors wrote better summaries (i.e., more detailed and better organized) and had a larger working memory capacity than the other groups. Hyönä and his colleagues reasoned that more cognitive resources might have contributed to integrating a wider range of text information during reading, which led to well-structured representations of topic structure in the post-reading task. The importance of cognitive resources was also observed in L2 research, although the focus was narrative reading. In Kato (2014), Japanese university students with larger working memory capacity better linked the protagonists' sub-goals with superordinate goals, compared to the participants with smaller working memory capacity.

Therefore, the present study considered L2 reading proficiency as a reader factor throughout the experiments. Based on the abovementioned findings of previous studies, L2 reading proficiency might affect reader understanding of topic structure in two ways. First, L2 reading proficiency seems to contribute to linking and organizing distant propositions in texts. Second, L2 reading proficiency seems to influence resource allocation for higher-level processes through lower-level processes. L2 research (Morishima, 2013; Yoshida, 2003) has demonstrated that low-proficiency readers are not likely to allocate sufficient cognitive resources to higher-level processes (e.g., inference generation) because many cognitive resources are consumed by lower-level processes (e.g., basic reading skills such as word recognition and syntactic parsing).

In addition to the reader factors, some past studies compared measurement tests, indicating that globally coherent comprehension might have differed among measurement tests. For example, Williams and her colleagues (Williams et al., 1984; Williams et al., 1981) adopted several tests to measure main idea comprehension of expository paragraphs: multiple-choice questions, writing one-sentence summaries, and adding one sentence to each paragraph. The results revealed that the multiple-choice questions (i.e., a receptive test) were easier than the other tasks (i.e., productive tests). Further, other studies also found differences between productive tests. Goldman et al. (1995) conducted recall and summary tasks with L1 readers, showing that the recall transcripts included more supporting details than the summary transcripts. A similar tendency has also been observed in L2 research. Riley and Lee (1996) revealed that the summary protocols included more main ideas than details, and vice versa in the recall transcripts. Although Kintsch and van Dijk's (1978) model assumes that reproduction tasks (e.g., recall and summary tasks) reflect readers' hierarchical mental representations, the summary transcripts represented the test takers' editing processes as well as text comprehension abilities.

As reviewed in these studies, each measurement task has its own characteristics, which might affect globally coherent comprehension. Hence, to examine reader understanding of topic structure from multiple perspectives, the present study will explore this research topic adopting different tests across the experiments. In the next section, measurements for reading comprehension will be overviewed.

2.3 Methodologies to Measure Coherent Text Comprehension

Previous studies have adopted various measurements to assess coherent comprehension of texts. Researchers use *on-line measures* to evaluate during-reading processes and *off-line* *measures* to analyze complete forms of mental representations after reading. Traditionally, the term on-line measures was used to refer to “tasks examining the process during its operation” (Swinney, 1979, p. 647), such as the reading time method. However, the present study will follow Jiang’s (2012) recommendation that response times collected immediately after reading can also serve as on-line measures for tasks (e.g., the lexical decision task, recognition task) that require fast or temporally constrained responses and measure response times. Such immediate and quick responses maximize the possibility for response times to reflect the ongoing cognitive processes under examination. On the other hand, off-line measures examine text memory after reading, such as the written recall task. The next section will review specific on-line and off-line measures adopted by previous studies to assess coherent comprehension.

The reading time method

Many previous studies have used self-paced reading times as an on-line measure. This method assumes that participants read texts at the same pace as their internal comprehension processes, and that reading times represent changes in the processing load (Haberlandt, 1994). Past research created conditions in accordance with targeted reading processes and compared reading times between the conditions. For example, Murray and McGlone (1997) and Hyönä and Lorch (2004) manipulated the explicitness of the major topics beyond paragraphs (i.e., topic overview, headings) and compared reading times for the topic sentences (i.e., the subtopics) initiating the subsequent paragraphs. The logic for interpreting the target reading times was as follows: If the participants understand topic structure during reading, they should comprehend the subtopics more easily and smoothly with the explicit major topics than without them.

The advantage of the reading time method is that researchers can measure ongoing cognitive processes at the time of reading. On the other hand, the disadvantage is that reading times do not completely specify what causes changes in processing (Altmann & Steedman, 1988; McKoon & Ratcliff, 1980) because reading times are determined by several reading subskills. In case of the example in the previous paragraph, the major source of the difference in reading times between conditions was probably topic structure processing. However, that difference might have reflected other reading processes such as lower-level processes (e.g., word recognition, syntactic parsing) or other types of inference (e.g., local inference, elaborative inference).

The decision method

As discussed above, the reading time method does not clarify the precise source of variations in reading times among the possible reading subskills. In this respect, the decision method has an advantage over the reading time method. That is, these tasks more directly encourage the activation of concepts by requiring participants to make yes or no responses to target probes (Haberlandt, 1994). The most frequently used decision tasks are the lexical decision task, which requires a respondent to determine whether the target word is a real word, and the recognition task, which requires a respondent to determine whether the target word appeared in the text.

One of the largest differences between these tasks is that the recognition task requires a respondent to refer to their text memory (McKoon & Ratcliff, 1984). Hence, as the recognition task better reflects text comprehension, this task might reflect task-induced processes as well as reading processes. Meanwhile, the lexical decision task does not require the participant to refer to their text memory, and correct word recognition does not necessarily require understanding of the text. Because of the above differences, the lexical decision task and the recognition task have both advantages and limitations.

The decision tasks are often used in the *priming paradigm* (e.g., Guindon & Kintsch, 1984; Meyer & Schvaneveldt, 1971). A priming paradigm is used to measure the connection between two concepts by presenting two stimuli in succession (Jiang, 2012). When two concepts are connected in a respondent's mind, presenting a preceding stimulus (i.e., *priming stimuli*) either promotes or suppresses their response to a subsequent stimulus (i.e., *target probe*); this is called the *priming effect*. McKoon and Ratcliff (1984) stated that response times to the recognition task are longer than other decision tasks, which reflect the priming effect more noticeably.

Similar to the priming paradigm, Lorch (1993) interrupted reading, presented one of the cues (i.e., either the paragraph subtopics or general information; corresponding to priming stimuli) before restarting reading, and measured the reading times for the first sentence after resumption (corresponding to target probes). Lorch aimed to investigate whether the L1 readers linked the supporting details with the paragraph topics and assumed the following: If they link the paragraph topics with the supporting details, they should read the sentences immediately after resumption more smoothly with the presentation of the paragraph topics than with general information alone.

In sum, the advantage of the decision tasks is that the tasks directly reflect the activation of concepts, requiring participants to respond in a certain manner. Moreover, the combination of the decision tasks with the priming paradigm allows researchers to measure coherent comprehension between activated concepts. On the other hand, the decision tasks, as well as the reading time method, do not reveal the specific contents of readers' thoughts during their cognitive processes. Finally, it should be noted that the lexical features of the target probes (e.g., word length, frequency, and familiarity) affect response times to the decision tasks (de Groot, Borgwaldt, Bos, & van den Eijnden, 2002). Hence, when conducting decision tasks, it is necessary to control lexical features of target probes so that factors other than research interest influence the performance.

The think-aloud method

The think-aloud method asks participants to verbally report their thoughts during reading. The verbal reports are transcribed and then parsed into clauses; each clause is in turn classified into categories such as surface analysis at word and sentence levels, in-text inference, reader response, rereading, self-monitoring, comment on text structure, and so on. Think-aloud studies often calculate the proportions of each think-aloud category to examine resource allocation during reading. In previous L2 research, it has been found that most think-aloud comments are dedicated to word and sentence analysis, which makes it difficult to allocate the necessary cognitive resources for higher-level processing (e.g., local and global inference, elaborative inference; Horiba, 2000, 2013; Kimura, 2015a; Shimizu, 2015; Yoshida, 2003).

An advantage of the think-aloud method is that it allows researchers to assess the contents of comprehension processes at specific points in time during reading (see Ericsson & Simon, 1993, for a review). The contents of reading processes are not obtained from the reading time method and decision tasks. On the other hand, a disadvantage is that think-aloud comments reflect the readers' conscious processes alone, and do not represent the automatic processes of reading comprehension. Moreover, verbalizing their thoughts might change readers' cognitive processes from a normal reading situation.

The written recall task

As shown above, the previous studies adopted on-line measures to evaluate coherence-building processes during reading. In addition, these studies also used off-line measures to examine whether readers represented coherent comprehension in their text memory. Among the off-line measures, the written recall task is one of the most widely adopted methods for studying text comprehension. The written recall task is a post-reading task that requires readers to write down as much as they can from their memory and understanding of the text without looking back.

For example, to measure reader understanding of the links between the major topics and paragraph subtopics, several recall studies manipulated the explicitness of the major topics in the texts, and then compared the participants' recall production of the subtopics (e.g., Hyönä & Lorch, 2004; Lorch & Lorch, 1985; Lorch et al., 2001). These studies assumed that if the participants could represent topic structure in their text memory, they should be able to recall propositions from more paragraphs when the major topics were explicit than when they were not.

Other studies adopted the cued recall task to examine the links between propositions (e.g., Lorch & Lorch, 1996; Rawson & Kintsch, 2004; Ushiro et al., 2007; Ushiro et al., 2014). The cued recall task is used to examine whether readers construct and retain connections between the propositions presented in cues and targeted propositions in the texts. When the readers represented content overlap between these propositions in their text memory, presenting the recall cues should improve the recall of targeted propositions based on that overlap (e.g., Wolfe, Magliano, & Larsen, 2005). For example, in Lorch and Lorch (1996), the participants were or were not presented with the major topics as cues in the recall task after reading expository texts with major topics and paragraph subtopics. They compared how many paragraphs their participants recalled propositions from using the recall cues, expecting that more paragraphs would be recalled with the cues than without them.

Some studies have reported that the recall cues were not effective, pointing out the possibility of cue redundancy (Rawson & Kintsch, 2004; Ushiro et al., 2007). For example, although Ushiro et al. (2007) provided important or detailed information as a recall cue, these cues failed to increase the participants' recalls. They reasoned that when readers construct very robust connections between the cue information and targeted propositions in the texts, the readers have spontaneous access to cue information retained in their text memory, regardless of whether they are provided with recall cues. However, even if cue redundancy appears in immediate recall, the cues might become non-redundant in delayed recall, improving their recall of the target propositions (Ushiro et al., 2014).

In addition to the quantitative comparison of recall production, the structure of recall protocols was also analyzed qualitatively in past studies as a reading comprehension measure. Meyer (1975) analyzed reader comprehension of the hierarchical structure of expository texts (i.e., content-structure analysis), and revealed that readers who identified and understood the hierarchical structures of the texts recalled more than those who did not Meyer & Freedle, 1984). This finding has also been replicated in L2 research (Carrell, 1984a, 1992; Ghaith & Harkouss, 2003).

As discussed above, previous studies have adopted various on-line and off-line measures to examine during-reading processes and readers' text memory. Because each measure has different characteristics, it is important to follow-up and supplement findings with different measurements. Jiang (2012) stated that "follow-up experiments help determine if a finding is an outcome of adopting a specific task or reflects a more general phenomenon" (p. 77). To capture an objective picture of L2 reader understanding of topic structure, the present study will use different measures in several experiments.

2.4 The Effects of Educational Interventions on Reading Comprehension

2.4.1 The Effects of Reading Instructions on Reading Comprehension

Previous research has sought to support reading comprehension through educational interventions. Among educational interventions, the effects of instructing participants to read texts for achieving a task, particularly in L1 reading, have been widely studied. These studies were based on the standards of coherence (e.g., van den Broek, Risden, et al., 1996) to examine the effects of reading instructions. That is, when a reading instruction is given, readers are assumed to adjust their standards of coherence accordingly, which determines the types and strength of coherence they should build in order to accomplish the given task. To meet the standards of coherence, readers alter their during-reading processes, which in turn affects their text memory.

Several studies have compared during-reading processes and after-reading memory, when L1 university students read expository texts for entertainment and study (e.g., Bohn-Gettler & Kendeou, 2014; Linderholm & van den Broek, 2002; van den Broek et al., 2001; Yeari, van den Broek, & Oudega, 2015). For example, Yeari et al. (2015) indicated that L1 readers took longer when reading for study purposes than for entertainment purposes, suggesting a more attentive mindset during strategic reading. Moreover, van den Broek et al. (2001) found that L1 readers strategically increased their coherence-building inferences and decreased their association and evaluation practices when reading for study purposes, in comparison to reading for entertainment purposes. Regarding the instruction effect on text memory, they also observed higher recall rates after reading for study than for entertainment. This was replicated by Bohn-Gettler and Kendeou (2014), who further reported that main ideas were recalled better after reading for study purposes. In their studies, the L1 readers strategically and more attentively engaged in deeper comprehension processes when reading for study purposes, which enhanced their text memory, including their memory of the main ideas.

Other studies provided more specific reading instructions, aiming to support globally coherent comprehension (e.g., Goldman et al., 1995; Lorch, Lemarié, & Chen, 2013; Lorch,

Lorch, & Mogan, 1987; van der Schoot, Horsley, & van Lieshout, 2010). In particular, Lorch and his colleagues (Lorch et al., 2013; Lorch et al., 1987) aimed to aid reader understanding of topic structure using outline instruction. They attempted to help readers link the major topics and subtopics by instructing them to read texts with the goal of writing outlines—that is, itemizing the major topics and subtopics using "simple bulleting with a two-level structure distinguishing major topics and their subtopics" (Lorch et al., 2013, p. 63). Because of this specific instruction, the outline instruction explicitly requires "identification of text topics and identification of the hierarchical relations among the text topics" (Lorch et al., 2012, p. 268). In particular, the latter cognitive process is characteristic of the outline task, but not other tasks (e.g., a summary task) similarly aimed at globally coherent comprehension. In the summary task, readers are asked to focus on important topics in texts and condense the texts to a specific length (e.g., Ushiro et al., 2009). However, the summary task does not direct readers' attention to the hierarchical relations among the topics. For example, some readers might identify individual topics in texts but fail to understand their hierarchical relations, which would lead to fragmented understanding. In other words, successful text comprehension requires not only the identification of the important topics but also the organization of these topics in a hierarchical structure. The outline instruction is assumed to require both cognitive processes by requiring readers to itemize the major topics and subtopics through bulleting with a two-level structure.

In Lorch et al. (2013), L1 university students read texts including headings or preview sentences that represented major topics and subtopics. They were instructed to read texts with the goal of comprehension in Experiment 1 and writing outlines in Experiment 2. In Experiment 1, more topics were recalled from the texts with headings that made the topic structure more salient visually than from the texts with preview sentences. In Experiment 2, however, the topics were recalled to a similar extent between both texts with headings and preview sentences. The results of the two experiments suggested that the outline instruction enhanced the readers' memory of links between the subtopics and major topics described in the preview sentences. Additionally, Lorch et al. (1987) investigated whether reading times were longer for major-topic shifts than subtopic shifts to assess reader understanding of topic structure. The difference between major-topic and subtopic shifts increased from reading for comprehension to reading for outlining, although this tendency was only observed among the more proficient readers. Based on these findings, it was suggested that the outline instructions helped the L1 readers understand topic structure during reading and in a post-reading task, and that its effects might differ among individual readers.

In addition to Lorch et al. (1987), several L1 studies have suggested that the effects of instructions can differ among individuals, although reading instructions were effective in most studies. For example, in Linderholm and van den Broek (2002), such positive effects of reading for study purposes were not found for L1 readers with less working memory in terms of reading processes (e.g., coherence-building inferences) and text memory. A similar result was obtained in Bohn-Gettler and Kendeou (2014). In their study, while L1 readers with more working memory decreased their non-coherence processes (e.g., association) from reading for entertainment to reading for study purposes, readers with less working memory failed to flexibly adjust their reading processes. Taken together, some deficits such as poor text comprehension or poor working memory might have prevented the readers from engaging in coherence-building processes and thus from encoding text comprehension in their memory.

However, the effects of reading instructions are inconsistent among L2 studies. Previous studies have demonstrated that reading instructions are effective for narrative text comprehension, compared with expository text comprehension (e.g., Nahatame, 2014). Regarding globally coherent comprehension, Ushiro et al. (2017) instructed Japanese university students to imagine the situations of the narrative texts while reading. They found that the image instruction improved the participants' maintenance of coherence between distant sentences. Moreover, Kimura (2012) instructed Japanese university students to read for comprehending themes throughout narrative texts, which contributed to better theme comprehension and better recall in the post-reading task.

On the other hand, the effects of reading instructions on expository text comprehension are complicated. Horiba (2000, Experiment 2) instructed L2 readers to freely read expository texts or pay attention to coherence between the current sentences and prior/later sentences. Reading for coherence did not influence think-aloud comments during reading but did enhance recall after reading. In more recent research, Horiba (2013, Experiment 2) gave Japanese university students the following three instructions: (a) pay attention to words and expressions in the text (i.e., the expression condition); (b) mentally visualize the events, places, and actions in a text (i.e., the image condition); and (c) evaluate the author's views in comparison with their own views (i.e., the critique condition). However, there were no notable differences among the instruction conditions in the think-aloud comments (coherence-building inference) or recall productions. In relation to reader understanding of hierarchical topic structure, Kimura (2014, 2015b) instructed Japanese university students to read expository texts to understand themes beyond paragraphs. The researcher found that the reading instruction failed to aid performance in theme comprehension or affect during-reading processes. The instruction effect on text memory was inconsistent: It increased recall production in Kimura (2015a) but not in Kimura (2014).

Thus, it has been found that reading instructions consistently assisted L1 readers' deep comprehension processes (e.g., coherence-building processes) and text memory, and that instruction effects are possibly affected by individual reader differences such as working memory and text comprehension skills. However, instruction effects have been inconsistent in L2 reading studies, especially in expository reading. Moreover, the instruction effects were inconsistent between on-line and off-line measures and across past studies. Even if instruction effects appeared, the effects did not significantly affect reading processes or text memory. There are two possible reasons for the weak effects of reading instructions on L2 reading comprehension. First, the L2 readers might have had difficulty allocating enough cognitive resources to adapt their cognitive processes, even when they were given reading instructions (e.g., Kimura, 2014, 2015a). L2 research has demonstrated that lower-level processes (e.g., word recognition, syntactic parsing) are not sufficiently proficient or automated in L2 readers, in comparison to L1 readers (e.g., Horiba, 2013; Morishima, 2013; Yoshida, 2003). Because this requires many cognitive resources, L2 readers can often not afford to engage in higher-level processes (e.g., coherence building) to achieve their standards of coherence or accomplish the given reading instructions. Another possibility is that simply giving reading instructions is not sufficient to direct L2 readers' cognitive processes toward deep and coherent comprehension. In the past research reviewed in this section, the participants did not actually engage in the tasks. Although L1 readers were able to flexibly alter their cognitive processes in response to the reading instructions alone (e.g., van den Broek et al., 2001), L2 readers had difficulty doing so (e.g., Horiba, 2013; Kimura, 2014). Thus, to orient L2 readers' cognitive processes, it might be necessary for them to actually engage in the tasks, rather than simply receive the reading instructions.

2.4.2 The Effects of Task Engagement on Reading Comprehension

Previous studies have also aimed to support reading comprehension through task engagement. These studies are based on two frameworks: the *cognitive load theory* (Sweller, 1988) and *activity theory* (e.g., Stull & Mayer, 2007). The cognitive load theory hypothesizes that readers need to engage in appropriate learning processing (i.e., generative processing) for deep text comprehension within the limits of their available cognitive resources. This theory assumes three cognitive processes during learning: (a) irrelevant processing to accomplishing a task (i.e., *extraneous processing*), (b) mentally representing a text (i.e., *essential* or *intrinsic processing*), and (c) deeper cognitive processing such as selecting and organizing relevant text information (i.e., *generative* or *germane processing*). On the other hand, activity theory hypothesizes that readers are not able to comprehend texts deeply with passive behavior such as simply reading. Rather, deep text comprehension is achieved when readers engage in productive activities or tasks, selecting relevant ideas from a text and organizing them into a coherent structure. However, when the given task is complicated, readers might be confused about how to accomplish the task. In this case, more cognitive resources are consumed in the extraneous processing of figuring out how to complete the task; thus, fewer resources are available for generative processing.

Empirical studies have explored the effects of task engagement on text comprehension mainly in L1 reading. To support globally coherent comprehension, past studies have focused on generating graphic organizers. For example, Ponce and Mayer (2012) compared eye movements when L1 university students were generating a graphic organizer during reading, taking notes during reading, or just rereading without engaging in a specific task (i.e., control condition). Compared with the other groups, the graphic organizer group showed more saccades across the major topics and better linked the major topics in the summary task, indicating that generating a graphic organizer supports reading processes and text memory at a global level. Moreover, fewer saccades between the text and task column were made in the graphic organizer group than in the note-taking group at the beginning of the experimental session. This revealed a behavioral difference between the groups: The former group generated a graphic organizer after reading over the entire text whereas the latter group simultaneously read the text and took notes.

On the other hand, in Stull and Mayer (2007, Experiments 2 and 3), the generation of graphic organizers by the participants at the time of reading was not effective, compared with the authors providing graphic organizers. The learner-generated group obtained lower scores for deep comprehension and took longer to finish, compared to the author-provided group. The researchers interpreted this finding as showing that generating graphic organizers was too resource-demanding and thus not effective for their participants. In this study, the task engagement might have increased the amount of extraneous processing among the participants and decreased their generative processing.

Other studies have found that generating graphic organizers became effective for children's reading comprehension after providing training on graphic organizers (Ciullo, Falcomata, & Vaughn, 2014; Chang, Sung, & Chen, 2002; Redford et al., 2012). Regarding globally coherent comprehension, in Merchie and Van Keer (2016), fifth and sixth graders were divided into the (a) learner-generated group, who generated graphic organizers during reading; (b) author-provided group, who were given graphic organizers; or (c) control group. Groups A and B created better graphic organizers; in particular, Group A outperformed Group B in their understanding of the hierarchy of the major topics and subtopics. Thus, creating graphic organizers during reading is helpful for L1 children, if combined with prior instruction.

In contrast to L1 research, there have been few L2 studies that have examined the effect of task engagement on text comprehension. Rather, L2 research has mainly explored the effects of reading instruction, as reviewed in Section 2.4.1. Of the limited studies, Yoshida (2012) divided Japanese EFL university students into the three following groups: (a) those who wrote an outline at the time of reading, (b) those who answered embedded questions, and (c) those who simply read without engaging in a specific task (i.e., control group). Quantitatively, their recall productions did not noticeably differ according to group or text genre (i.e., narrative vs. expository texts). However, their recall quality showed that the outline group linked the main events and subordinate information in the recall of the narrative text. Yoshida did not find any notable effects of task engagement on L2 discourse comprehension, although many L1 studies have demonstrated the positive effects of task engagement (e.g., Merchie & Van Keer, 2016). Yoshida (2012) reasoned that the participants' L2 linguistic proficiency level was mismatched with the task demands. Specifically, her participants' lower-level processing abilities were insufficient, which consumed much of the cognitive resources needed for the higher-level processes that contribute to deep comprehension.

In sum, previous research has shown that task engagement often supports L1 readers' text comprehension, including globally coherent comprehension (e.g., Merchie & Van Keer, 2016). As activity theory hypothesizes, engagement in a productive task challenges readers to think deeper and engage in processes such as the selection and organization of relevant text information. On the other hand, task engagement has been shown to support L2 reading comprehension to some extent (Yoshida, 2012). In the case of L2 reading, if the readers' L2 proficiency level and the task demands are mismatched, task engagement will not occur. This suggests the importance of selecting appropriate tasks for students to effectively support their reading comprehension processes.

2.5 Links to the Present Study

As discussed in this chapter, successful text comprehension requires a reader to build a globally coherent mental representation throughout a text. To this end, readers need to not only understand important topics representing sections, but also to establish hierarchical links between subtopics and major topics of larger sections of texts (e.g., van Dijk & Kintsch, 1983). Although many studies have investigated coherence-building processes and text memory, there are two major limitations of the literature that signal the need for further exploration of this topic.

First, preceding studies have not sufficiently tackled the linking of subtopics and major topics across paragraphs in L2 expository reading. Whereas previous studies have mainly examined building global coherence in narrative reading, research in expository reading is still limited. Of the limited studies, it has been indicated that L1 and L2 readers understood the hierarchical relations that linked the explicit/implicit paragraph subtopics with supporting details during reading and in post-reading tasks (e.g., Goldman et al., 1995; Guindon & Kintsch, 1984; Mori, 2015; Ushiro et al., 2008). Moreover, past research has found that L1 and L2 readers had difficulty inferring implicit major topics by integrating paragraph subtopics (e.g., Brown & Day, 1983; Kimura, 2015b). However, few L2 studies have sufficiently examined how readers link explicit major topics and subtopics. Although it was difficult for readers to independently infer implicit major topics across paragraphs, they might have less difficulty linking explicit major topics and subtopics written in texts. In addition to the limited L2 research, L1 studies on linking explicit major topics and subtopics also have limitations. In particular, post-reading tasks in these studies did not fully focus on reader memory of the specific links between major topics and subtopics (e.g., Hyönä & Lorch, 2004). That is, although these studies investigated reader understanding of the links between major topics throughout texts and paragraphs, the studies did not distinguish the paragraph subtopics from the supporting details. To better understand this topic, the present study focused on the specific relations between the major topics and subtopics.

In addition, few L2 studies have examined the effects of educational interventions that directly aim to support reader understanding of hierarchical topic structure. As for reading instructions, in L1 research, Lorch and his colleagues (Lorch et al., 2013; Lorch et al., 1987) demonstrated the positive effects of giving outline instructions, which explicitly required the participants to select and organize relevant information to fit into hierarchical structure. In L2 research, although Kimura (2012, 2014, 2015b) aimed to support globally coherent comprehension by instructing her participants to focus on themes during reading, this instruction did not specifically require the participants to organize the topics into different hierarchical levels. Furthermore, whereas L2 studies have attempted to intervene in reader understanding of topic structure through task engagement (e.g., Merchie & Van Keer, 2016), few L2 studies have been conducted for that specific purpose. Although Yoshida (2012) had her participants write an outline during reading, the effect of this task was not significant; it simply showed the tendency to link main events and subordinate information in a narrative text. However, because her study used a time limitation in the experiment and the participants were told to prioritize reading over outlining, the readers might not have fully engaged in the outlining task. To capture a global picture of educational interventions aimed at reader understanding of topic structure, more studies are needed.

To fill the gaps in the existing literature, the present study conducted a total of five experiments, as follows. Study 1 included three experiments (i.e., Experiments 1 to 2B) to examine whether the Japanese EFL readers linked the major topics throughout the texts with the paragraph subtopics during expository reading and in a post-reading task. Experiment 1 adopted the cued recall task to examine whether Japanese EFL readers linked major topics with subtopics in text memory. Moreover, Experiments 2A and 2B used the priming paradigm to investigate whether Japanese EFL readers linked the major topics with subtopics during reading. In these experiments, to compare globally coherent comprehension at paragraph and text levels, reader understanding of the links between paragraph subtopics and supporting details was also examined.

Study 2 involved two experiments (i.e., Experiments 3 and 4) to help Japanese EFL readers link the major topics with subtopics during reading and in post-reading tasks. Experiment 3 examined whether giving the outline instructions was effective for topic structure processing and its memory. Additionally, Experiment 4 investigated whether Japanese EFL readers better understood topic structure during reading and represented it in their text memory by having them write an outline at the time of reading. Experiments 3 and 4 adopted the outline task as educational interventions that directly and explicitly oriented readers into understandings of topic structure, including selecting and organizing relevant information (i.e., major topics, subtopics). In Experiment 4, there were no time limits on reading and outlining.

The present study examined topic structure processing during reading, its memory in post-reading tasks, and the effects of educational interventions on these elements. This dissertation discussed how and when Japanese EFL readers best understood topic structure, and the effects of two educational interventions on both on-line reading processes and off-line text memory.

Chapter 3
Study 1: On-Line Processing and Off-Line Memory of Topic Structure

3.1 Experiment 1: Off-Line Memory of Topic Structure

3.1.1 Purpose and Research Question

The purpose of Experiment 1 was to examine whether Japanese EFL readers can represent the topic structure of expository texts in their text memory. While previous studies have demonstrated that readers can represent links of subtopics summarizing a paragraph and their supporting details in their text memory (e.g., Ushiro et al., 2008), few studies have explored readers' memory of links between subtopics and major topics summarizing larger units of texts. Rather, these studies demonstrated that a subtopic was represented in text memory if just a single piece of information within the unit of the text was recalled regardless of its importance (e.g., Lorch & Lorch, 1985; Lorch et al., 2001).

Therefore, the present study examined whether Japanese EFL reader could link subtopics of paragraphs with the major topic of the text. Moreover, to compare coherence beyond paragraphs with coherence within a paragraph, this study also examined students' memory of links between the subtopics and their supporting details. To measure two kinds of links represented in text memory, a cued recall task was adapted from previous studies (e.g., Ushiro et al., 2007). Specifically, the recall rates of subtopics were compared among three conditions: (a) A major topic was presented as a recall cue (i.e., *major cue condition*), (b) a supporting detail was presented as a recall cue (i.e., *detail cue condition*), and (c) no recall cue was presented as a baseline (i.e., *control condition*). A recall cue improves access to linked information represented in the reader's text memory when a content overlap is constructed between the cue information and the text information (e.g., Rawson & Kintsch, 2004; Wolfe et al., 2005). Hence, if the participants linked the subtopics with the major topics and supporting details, the major and detail cues were expected to increase the subtopic recall rates, compared with the control condition.

In addition, cue effects on the total recall rates were also investigated, as well as the subtopic recall rates, to compare the result with that of Ushiro et al. (2007). Their study presented subtopics of a paragraph and supporting details as recall cues, revealing that neither of the cues increased the total recall rates, compared with the control condition without a recall cue. Ushiro et al. (2007) suggested that this was caused by cue redundancy (Rawson & Kintsch, 2004; Ushiro et al., 2014). That is, the participants constructed well-structured representations among pieces of text information, recalling much text information regardless of the conditions (i.e., the cue effects might have been overwhelmed by sufficient text comprehension). However, because the cued recall task in Experiment 1 targeted the specific links between major topics and subtopics (the more important parts of topic structure), the cue effects might have been more noticeable. Therefore, the hypothesis and the research question (RQ) in Experiment 1 is as follows:

RQ1: Can Japanese EFL readers represent topic structure in their text memory?

3.1.2 Method

3.1.2.1 Participants

Seventy Japanese EFL undergraduate students and graduate students participated in the immediate session, and 40 of them also participated in the delayed session. Their majors were education, health and physical education, and humanities and culture. All the participants were native speakers of Japanese who had studied EFL for at least six years in Japanese formal education. None of them participated in the other experiments of this research.

Their self-reported scores on the standardized tests were as follows: the TOEIC listening and reading test (range: 525–668; 501–600: $n = 1$; 601–700: $n = 1$), the TOEFL ITP test (433 to 580; 401–500: $n = 5$; 501–600: $n = 3$), and EIKEN grades (3rd to pre-1st; 3rd: $n = 8$; pre-2nd: $n = 8$; 2nd: $n = 21$; pre-1st: $n = 2$). Based on the alignment studies (Dunlea, n.d.; Educational Testing Service [ETS], 2015, 2017) of the Common European Framework of Reference for Languages (CEFR), participants' general English proficiency was estimated to be beginner to upper-intermediate (i.e., about levels A1 to B2 on the CEFR). It should be noted that some participants reported none of their TOEIC, TOEFL ITP, or EIKEN scores, whereas other participants reported all of them.

3.1.2.2 Materials

English-reading proficiency test

An English-reading proficiency test was created to measure the participants' abilities to understand English discourses (see Appendix 1). Although L2 reading proficiency includes various subskills such as vocabulary and grammatical knowledge, discourse-comprehension skills were measured in all the experiments of the present thesis. This is because the focus of the present study was comprehension of topic structure, which is likely to be related to discourse-comprehension skills.

To measure L2 discourse comprehension, test items were collected from retired copies of the reading subsection of the EIKEN test (Gakken, 2005; Obunsha, 2009; Yamada, 2003). Although other standardized English tests (e.g., TOEFL, TOEIC) are also popular in Japan, the EIKEN test has been conducted all over Japan (more than 2 million examinees annually at 18,000 test sites) and for a long time (i.e., over 50 years) to measure the English proficiency of Japanese EFL learners. Each of the seven grades of the EIKEN test has different sets (with different difficulties) but with a similar test format. Because many Japanese EFL learners have taken the EIKEN test and are thus used to its format, the present thesis adopted the EIKEN test as an L2 reading proficiency test for the Japanese EFL learners.

The English-reading proficiency test consisted of four texts with 20 test items from the second grade (i.e., each text was paired with five items) and two texts with eight items from the pre-first grade (i.e., each text was paired with four items). The test items were all multiple-choice questions, each consisting of a correct answer and three distractors. Table 3.1 shows the length and readability of the texts used in the proficiency test.

Table 3.1

Length and Readability of the Texts in the English-Reading Proficiency Test

Text	Grade	Word	FKGL
Traffic trouble	Second	366	9.1
The blue-blooded crab	Second	341	8.4
Digital witness	Second	313	10.3
Spider silk	Second	362	9.8
Moon tales	Pre-first	477	12.5
Teaching nomads to read	Pre-first	485	11.2

Note. FKGL (Flesch-Kincaid grade level) was provided by Microsoft Word 2010's readability measurement tools.

Experimental texts

Six expository texts that represented topic structure were collected from previous studies (Carrell, 1992; Coté et al., 1998; Kintsch, 1990; Kobayashi, 2002; Lorch, 1993; Taylor & Samuels, 1983). Table 3.2 overviews length and readability of the texts, and Table 3.3 exemplifies a sample text (also see Appendix 2). Because this study focused on discourse-level comprehension, low-frequency words at Level 5 or above of the JACET 8000 word list (Japan Association of College English Teachers [JACET], 2003) were replaced with high-frequency words at Level 4 or below. In addition, the texts were revised to have similar topic structures, each including one major topic (i.e., a sentence summarizing the whole text), three subtopics (i.e., a sentence summarizing each paragraph), and supporting details (i.e., other sentences that explained the major topic and the subtopics in detail). Because units in English expository texts often start with an important sentence, each text began with the major topic and each paragraph with the subtopic, except for the first paragraph, which started with the major topic followed by the subtopic. In each text, the major topic was stated first and explained from three perspectives in the following paragraphs. To confirm the validity of the topic structure of the texts, two raters who majored in English education identified the major topics and subtopics. The inter-rater agreement rate was 96.88%. All disagreements were resolved through discussion.

Table 3.2

Length and Readability of the Experimental Texts in Experiment 1

Text	Words	FKGL	Overview of the texts
Argentina	152	7.9	Unique points of Argentina
Distance	153	7.4	Elements of measuring distance
Energy	150	8.8	Major problems of energy production
Environment	151	7.6	Environment rules of the United States
Support	167	7.2	Problems of supporting developing countries
Three Mile island	151	8.9	Effects of Three Mile island accident

Note. FKGL was provided by Microsoft Word 2010's readability measurement tools.

Table 3.3

A Sample Text Used in Experiment 1

There are elements for measuring distance. Economic distance is changed by the cost of movement from one place to another. Money and energy are related to any movement. Sending something by water is usually less expensive than sending over land. This holds true even when land routes are shorter.

Distance can be measured on the basis of time. Some maps use travel time instead of mile signals. This is because the measuring unit influences the usual relations among locations. It may take the same time to go from a single point to a location 10 miles north as going to a location 30 miles south.

Distance measuring varies with individual feelings. What may seem like a long trip to some individuals may seem short to other people. Even the same route going and coming can seem different to a single traveler. It depends on whether road conditions are good, and whether the trip is near the end.

Note. The major topic is boldfaced and the subtopics were underlined.

Three text sets, each consisting of two texts, were created and allotted to the three conditions of the recall task. Whereas the major topics and supporting details were presented as recall cues in the experimental conditions (i.e., *major cue condition* and *detail cue condition*, respectively), no recall cue was presented in the *control condition* as a baseline (Ushiro et al., 2007). By comparing the recall rates of the subtopics in the experimental conditions with those in the control condition, it was examined whether the participants linked the subtopics with the major topics and supporting details in their text memory.

Two texts were assigned to each text set. If only one text was assigned to each condition, it would be difficult to determine whether differences among the conditions were caused by the different recall cues or the different texts. Accordingly, the *Environment* text and the *Support* text were assigned to the major cue condition, the *Distance* text and the *Argentina* text to the detail cue condition, and the *Energy* text and the *Three Mile Island* text to the control condition. Hence, the major topics in the *Environment* and *Support* texts and the supporting details in the *Distance* and *Argentina* texts were translated into Japanese as the recall cues. The recall cues were presented in the participants' native language (i.e., Japanese) to prevent them from answering the task based on surface memory of English expressions, rather than on the contents represented in the cues. Table 3.4 shows the major and detail cues. Finally, six booklets were created to counterbalance the order of the three text sets (i.e., three conditions of the recall task) and the order of the two texts in each text set.

Table 3.4

Major and Detail Cues Presented in the Recall Task in Experiment 1

Condition	Text	Cue
Major	Environment	アメリカには環境被害に対する規則がある
	Support	発展途上国への支援は深刻な問題を引き起こしている
Detail	Argentina	カウボーイは農場でたくさんの牛の世話をしている
	Distance	陸路で輸送するより海路で輸送する方が安い

3.1.2.3 Procedure

Immediate session

Figure 3.1 illustrates the procedure of the immediate and delayed sessions in Experiment 1. The participants were tested in a group in both sessions. The immediate session was a single session that lasted for 75 minutes. At the beginning, the experimenter gave oral and written explanations of the experiment's purpose and procedure and then obtained the informed consent of the participants. After that, the participants reported their scores or grades of the popular standardized tests (i.e., EIKEN test, TOEIC listening and reading, TOEFL ITP) to describe their general English proficiency.

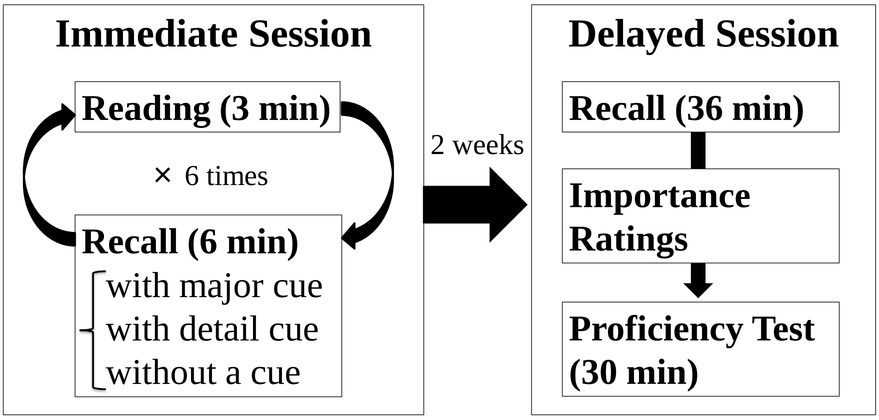

Figure 3.1. Procedure of Experiment 1.

In the reading session, one of the six counterbalanced booklets was given to the participants. Each text was presented on a page of the booklet, and the participants were instructed to read each text for comprehension within 2 minutes to answer the written recall task. In other words, the participants were notified of the recall task in advance. After reading each text, the participants were instructed to write in their L1 (i.e., Japanese) as much as they could remember in the form of discourse (i.e., not just itemizing words and phrases). The recall task was conducted after 6 minutes were spent reading each text. Before the main experiment, a pilot study was conducted with 12 Japanese university students, and it was confirmed that the time for reading and the recall task was sufficient for Japanese university students.

The participants answered the recall task of each text set (i.e., two texts) in one of the following conditions: (a) The major topic was given as a recall cue (i.e., the major condition),

(b) a supporting detail was given (i.e., the detail condition), or (c) no recall cue was given (i.e., control condition). The reading session and the written recall task were repeated for the six texts in the three text sets. After finishing the reading and recall of the half texts (i.e., three texts), the participants had a 5-minute break. Moreover, at the end of the immediate session, the participants were recruited for the delayed session. At that time, they were not notified of the delayed recall task.

Delayed session

The delayed session was conducted two weeks after the immediate session. The date of the delayed session was determined so that as many participants could participate as possible. The delayed session was a single session that lasted approximately 75 minutes. The delayed session included (a) the delayed recall task, (b) the importance rating task, and (c) the English-reading proficiency test.

First, the participants in the delayed session completed the delayed recall task. They recalled what they remembered in the same conditions as in the immediate recall task. The participants were allowed to start answering from any text they remembered. Because 6 minutes were given for the immediate recall task of each of the six texts, a total of 36 minutes were given for the delayed recall task, although most of the participants finished before the time limit.

After the delayed recall task, the participants were asked to rate the importance of each sentence ("How important each sentence is to understand the whole text"), using a 5-point Likert scale (1 = *not important at all*, 2 = *not important*, 3 = *neither unimportant nor important*, 4 = *important*, 5 = *very important*). This task was conducted to confirm the validity of the topic structures of the texts; to this end, it was confirmed whether the participants correctly identified the major topics, subtopics, and supporting details, as the two graduate students majoring in English education determined. Although a time limit was not set in the importance rating task, most of the participants finished it in around 10 to 15 minutes. After the importance rating task, the participants had a 5-minute break. Finally, they completed the English-reading proficiency test within 30 minutes. They were permitted to start with any questions that they could answer, and were instructed to answer as many questions as possible.

3.1.2.4 Scoring

Before scoring the recall protocols, two graduate students who majored in English education divided the experimental texts into idea units (IUs), based on Ikeno's (1996) criteria. The inter-rater agreement rate as the scoring reliability was 99.07%. All disagreements were resolved through discussion.

After the IU division, two graduate students majoring in English education scored 30% of the recall protocols (i.e., the immediate recall protocols from 19 participants and the delayed recall protocols from 13 participants). When two-thirds of the content of an IU was produced, one point was given; otherwise, no point was given. The inter-rater agreement rate as the scoring reliability was 92.09% for the immediate recall task and 93.47% for the delayed recall task. All disagreements were resolved through discussion, and the remaining 70% of the protocols were scored by one of the raters based on the criteria determined in the discussion. In each of the cue conditions (i.e., text sets), the recall rates of the subtopics and the total recall rates were calculated. When the total recall rates were calculated, the major topics and supporting details presented as the recall cues were removed.

3.1.2.5 Analysis

Thirty of the participants who only participated in the immediate session were excluded from the following statistical analyses. First, the importance ratings were compared among the major topics, subtopics, and supporting details to confirm the validity of the topic structure of the texts. That is, it was examined whether the participants rated the major topics more importantly and the supporting details less importantly than the subtopics. For that purpose, a 2 (proficiency: higher, lower) × 3 (information: major topic, subtopic, detail) two-way mixed analysis of variance (ANOVA) was performed for the importance ratings, with proficiency as a between-participant factor and information as a within-participant factor.

To answer the RQ, the subtopic recall rates were compared among the three cue conditions. To this end, a 2 (proficiency: higher, lower) × 3 (cue: major, detail, control) × 2 (recall: immediate, delayed) three-way mixed ANOVA was carried out for the subtopic recall rates, with proficiency as a between-participant factor, and cue and recall as within-participant factors. Moreover, an ANOVA of the same factorial design was carried out for the total recall rates to compare the result with the previous studies that did not focus on recall rates of subtopics among text information (e.g., Ushiro et al., 2007). Because the number of IUs was different among the texts, the recall rates were normalized by arcsine transformation prior to conducting the three-way ANOVAs.

3.1.3 Results

3.1.3.1 English-Reading Proficiency Test

Table 3.5 shows the descriptive statistics of the English-reading proficiency test (Cronbach's $\alpha = .70$). Forty of the participants answered the test in the delayed session and were divided into a higher-proficiency group ($n = 21$) and lower-proficiency group ($n = 19$) based on the median of the test scores. A t test confirmed that there was a significant difference in the test scores between the two groups, $t(38) = 8.54$, $p < .001$, $d = 2.70$.

Table 3.5

Descriptive Statistics of the English-Reading Proficiency Test in Experiment 1

		Second (k = 20)		Pre-first (k = 8)		Total (k = 28)	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	21	14.90	2.05	1.76	1.79	16.67	2.50
Lower	19	9.84	1.89	0.63	1.16	10.47	2.04

The English-reading proficiency of the higher-proficiency group was estimated to be between the second and pre-first grades of the EIKEN test (levels B1 and B2 of the CEFR; Dunlea, n.d.). The EIKEN test items were collected from retired copies before 2015, when test takers needed to score approximately more than 60% and 70% to pass the tests of the second and pre-first grades, respectively (EIKEN, 2016). The correct answer rates of the higher-proficiency group were higher than 60% (12.00 correct answers) for the second grade test but lower than 70% (5.60 correct answers) for the pre-first grade test. On the other hand, the English-reading proficiency of the lower-proficiency group was estimated to be below the second grade (lower than the CEFR level B1 level; Dunlea, n.d.). The correct answer rates of this group were lower than 70% for the pre-first grade test and 60% for the second grade test.

3.1.3.2 Importance Ratings

Table 3.6 and Figure 3.2 show the descriptive statistics of the importance ratings. Before examining the participants' comprehension of hierarchical links among the text information, it was confirmed whether the participants identified the hierarchical structure of the texts as the two graduate students determined. A 2 (proficiency: higher, lower) × 3 (information: major topics, subtopics, details) two-way mixed ANOVA was conducted for the importance ratings. Although neither the Proficiency × Information interaction, $F(1.49, 56.61)$

= 1.71, p = .196, η_p^2 = .04, nor the main effect of proficiency was found to be significant, F(1, 38) = 0.11, p = .741, η_p^2 < .01, the main effect of information was found to be significant, F(1.49, 56.61) = 62.47, p < .001, η_p^2 = .62. Multiple comparisons revealed that the importance ratings were significantly higher for the major topics than for the subtopics, M_{diff} = 0.61, p < .001, 95% CI [0.34, 0.87]; in turn, the ratings were significantly higher for the subtopics than for the supporting details, M_{diff} = 0.67, p < .001, 95% CI [0.45, 0.89]. The above results indicate that the participants correctly identified the major topics, subtopics, and supporting details in accordance with the hierarchy of the texts. In other words, the validity of the texts was confirmed. The results of the two-way ANOVA are summarized in Table 3.7.

Table 3.6

Descriptive Statistics of the Importance Ratings in Experiment 1

		Major topic		Subtopic		Detail	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	21	4.12	0.85	3.66	0.48	3.05	0.50
Lower	19	4.32	0.51	3.57	0.36	2.83	0.32

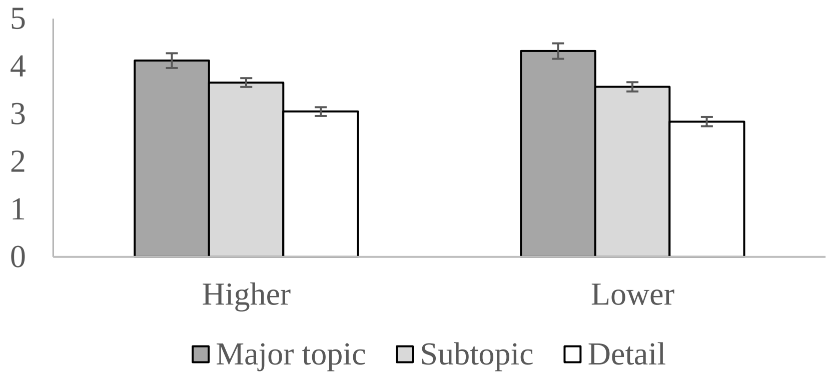

Figure 3.2. Mean importance ratings (± *SEM* bars) in Experiment 1.

Table 3.7

Summary of the Two-Way ANOVA for the Importance Ratings in Experiment 1

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.04	1.00	0.04	0.11	.741	< .01
Error (P)	13.05	38.00	0.34			
			Within-participants			
Information (I)	32.46	1.49	21.79	62.47	< .001	.62
I × P	0.89	1.49	0.60	1.71	.196	.04
Error (I)	19.74	56.61	0.35			

3.1.3.3 Subtopic Recall Rates

Table 3.8 and Figure 3.3 show the descriptive statistics of the subtopic recall rates. To address the RQ, a 2 (proficiency: higher, lower) × 3 (cue: major, detail, control) × 2 (recall: immediate, delayed) three-way mixed ANOVA was conducted. The results demonstrate that none of the interaction was significant (all *p*s > .10). On the other hand, the results revealed that main effects for proficiency, cue, and recall were significant (all *p*s < .05). Regarding the main effect of cue, multiple comparisons showed that the subtopic recall rates were significantly higher in the major cue condition than in the detail cue condition, $p = .004$, $M_{\text{diff}} = 11.00$, 95% CI [3.06, 18.94], and the subtopic recall rates were significantly higher in the detail condition than in the control condition, $p < .001$, $M_{\text{diff}} = 12.02$, 95% CI [7.66, 16.38]. This indicates that the recall cues, particularly the major cues, helped the participants recall the subtopics, compared to the control condition. Table 3.9 summarizes the results of the three-way ANOVA.

Table 3.8

*Descriptive Statistics of the Subtopic Recall Rates Normalized by Arcsine Transformation in Experiment 1 (*N *= 40)*

		Major cue		Detail cue		Control	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
				Immediate recall			
Higher	21	57.70	16.38	45.27	23.27	32.76	12.61
Lower	19	45.14	23.58	35.08	17.01	28.90	19.94
				Delayed recall			
Higher	21	33.84	14.62	22.84	17.16	6.00	12.41
Lower	19	24.07	18.46	13.57	14.42	1.02	4.47

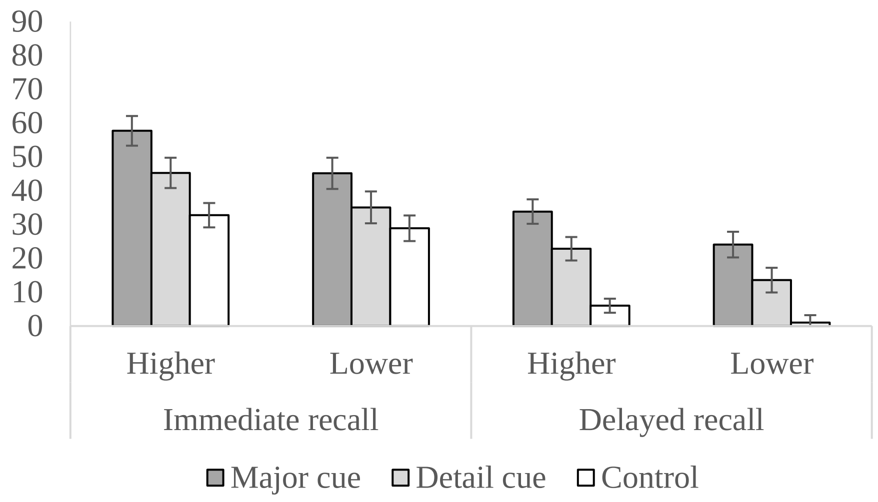

Figure 3.3. Mean of the subtopic recall rates (± *SEM* bars) normalized by arcsine transformation in Experiment 1 ($N = 40$).

Table 3.9

*Summary of the Three-Way ANOVA for the Subtopic Recall Rates in Experiment 1 (*N *= 40)*

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	4262.89	1.00	4262.89	7.55	.009	.17
Error (P)	21468.68	38.00	564.97			
			Within-participants			
Cue (C)	21149.37	1.48	14327.45	38.42	< .001	.50
C × P	503.57	2.00	251.79	0.92	.405	.02
Error (C)	20919.15	56.09	372.93			
Recall (R)	34230.60	1.00	34230.60	135.41	< .001	.78
R × P	11.30	1.00	11.30	0.05	.834	< .01
Error (R)	9606.23	38.00	252.80			
C × R	348.26	2.00	174.13	1.00	.373	.03
C × R × P	38.23	2.00	19.11	0.11	.896	< .01
Error (C × R)	13232.64	76.00	174.11			

Furthermore, a one-way repeated ANOVA was performed for the subtopic recall rates produced by all the participants (n = 70) in the immediate recall, with cue treated as a within-participant factor. This is because the three-way ANOVA revealed that the cue effects did not interact with proficiency, and data from many (30/70) participants were removed from the three-way ANOVA. The one-way ANOVA yielded a significant effect for cue, $F(1.83, 126.14) = 17.78$, $p < .001$, $\eta_p^2 = .21$. Multiple comparisons indicated that the subtopic recall rates were significantly higher in the major cue condition (M = 44.64, SD = 21.13) than in the detail cue condition (M = 35.82, SD = 19.98), p = .017, M_{diff} = 8.82, 95% CI [1.27, 16.38], and that the subtopic recall rates were significantly higher in the detail condition than in the control condition (M = 28.02, SD = 16.31), p = .004, M_{diff} = 7.80, 95% CI [2.06, 13.54]. The results of the one-way ANOVA revealed a similar tendency as the results of the three-way ANOVA. Table 3.10 summarizes the results of the one-way ANOVA.

Table 3.10

*Summary of the One-Way ANOVA for the Subtopic Recall Rates in the Immediate Recall Test in Experiment 1 (*N *= 70)*

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Cue (C)	9680.79	1.83	5295.58	17.78	< .001	.21
Error (Individual difference)	39137.26	69.00	567.21			
Error (C)	37576.21	126.14	297.90			

3.1.3.4 Total Recall Rates

Table 3.11 and Figure 3.4 show the descriptive statistics of the total recall rates (n = 40). A 2 (proficiency: higher, lower) × 3 (cue: major, detail, control) × 2 (recall: immediate, delayed) three-way mixed ANOVA was conducted. The three-way ANOVA showed a significant main effect of proficiency of $F(1, 38) = 4.41$, $p = .042$, $\eta_p^2 = .10$, indicating that the higher-proficiency group recalled more text information than the lower-proficiency group. Moreover, a significant main effect of cue, $F(2, 76) = 34.17$, $p < .001$, $\eta_p^2 = .47$, and a significant main effect of recall were observed, $F(1, 38) = 266.70$, $p < .001$, $\eta_p^2 = .88$. More importantly, a Cue × Recall interaction was also significant, $F(2, 76) = 28.65$, $p < .001$, $\eta_p^2 = .43$, whereas the other interactions were not significant (all ps > .10).

Post-hoc tests were performed to interpret the Cue × Recall interaction. The results of these tests showed a significant simple main effect of recall in all the recall conditions: F(1,

39) = 187.13, p < .001, η_p^2 = .83 for the major cue condition, F(1, 39) = 163.03, p < .001, η_p^2 = .81 for the detail cue condition, and F(1, 39) = 196.73, p < .001, η_p^2 = .84 for the control condition. Because the delayed recall task was conducted 2 weeks after the immediate recall task, the participants' text memory faded regardless of the cue conditions, which probably caused the significant main effect of recall. In addition, a simple main effect of cue was tested in the immediate recall and the delayed recall. The results showed that the simple main effect of cue did not reach significance in the immediate recall, F(2, 78) = 2.04, p = .137, η_p^2 = .05, whereas this simple main effect was significant in the delayed recall, F(2, 78) = 51.87, p < .001, η_p^2 = .57. Multiple comparisons demonstrated that the total recall rates were significantly higher in the major cue condition than in the detail cue condition, p < .001, M_{diff} = 8.51, 95% CI [5.16, 11.87]; in turn, the total recall rates were significantly higher in the detail cue condition than in the control condition, p < .001, M_{diff} = 5.93, 95% CI [2.65, 9.20].

Table 3.11

*Descriptive Statistics of the Total Recall Rates Normalized by Arcsine Transformation in Experiment 1 (*N*= 40)*

		Major cue		Detail cue		Control	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
				Immediate recall			
Higher	21	38.07	6.22	36.97	8.71	36.95	8.89
Lower	19	34.35	8.89	32.14	12.91	31.00	13.00
				Delayed recall			
Higher	21	25.95	5.42	16.40	9.94	10.33	11.24
Lower	19	19.12	9.76	11.75	10.49	5.99	9.23

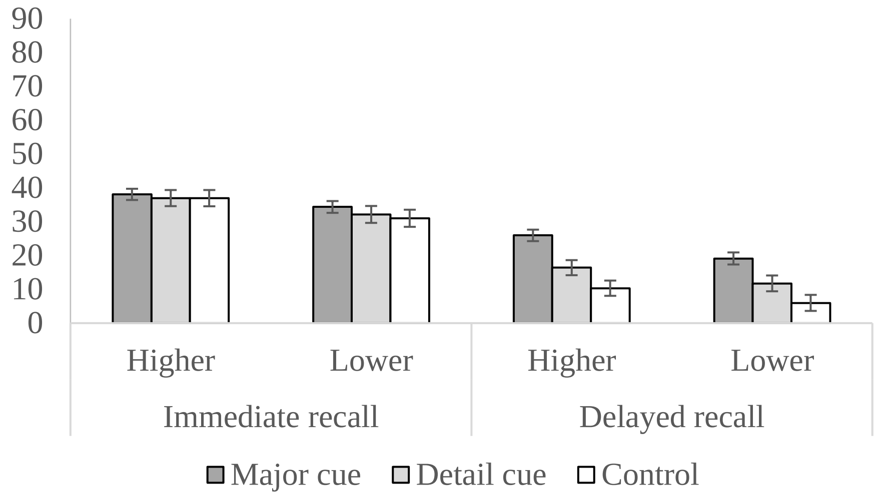

Figure 3.4. Mean of the total recall rates (± *SEM* bars) normalized by arcsine transformation in Experiment 1 ($N = 40$).

The results of the immediate recall suggested that the participants recalled their text memory well whether the recall cues were given and regardless of which ones were given. On the other hand, in the delayed recall, the participants recalled more text information when the recall cues were provided, in contrast to when they were not. The major cues prompted the participants' recall better, compared with the detail cues. The significant main effect of the cues was likely caused by the different cue effects on the total recall rates in the delayed recall. Table 3.12 summarizes the results of the three-way ANOVA for the total recall rates.

Table 3.12

*Summary of the Three-Way ANOVA for the Total Recall Rates in Experiment 1 (*N*= 40)*

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	1528.23	1	1528.23	4.41	.042	.10
Error (P)	13178.82	38	346.81			
			Within-participants			
Cue (C)	2795.09	2	1397.55	34.17	< .001	.47
C × P	3.06	2	1.53	0.04	.963	< .01
Error (C)	3108.81	76	40.91			
Recall (R)	23913.89	1	23913.87	266.70	< .001	.88
R × P	2.86	1	2.86	0.03	.859	< .01
Error (R)	3407.35	38	89.67			
C × R	1477.38	2	738.69	28.65	< .001	.43
C × R × P	58.30	2	29.15	1.13	.328	.03
Error (C × R)	1959.64	76	25.79			

Because the cue effect did not interact with proficiency similarly to the subtopic recall rates, a one-way repeated ANOVA was carried out for the total recall rates produced by all the participants (n = 70) in the immediate recall test, with cue as a within-participant factor. The results did not yield a significant effect for cue, $F(2, 138) = 2.20$, $p = .114$, $\eta_p^2 = .03$, showing that the total recall rates were not different among the major cue condition (M = 32.90, SD = 9.54), detail cue condition (M = 31.65, SD = 11.08), and control condition (M = 31.03, SD = 11.19). This supports the results of the three-way ANOVA, in which the simple main effect of cue was not significant in the immediate recall. Table 3.13 summarizes the results of the one-way repeated ANOVA.

Table 3.13

*Summary of the One-Way ANOVA for the Total Recall Rates in the Immediate Recall Test in Experiment 1 (*N*= 70)*

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Cue (C)	126.42	2	63.21	2.20	.114	.03
Error (Individual difference)	19430.31	69	281.60			
Error (C)	3958.20	138	28.68			

3.1.4 Discussion

The subtopic recall rates were the highest in the major cue condition, the second highest in the detail cue condition, and the lowest in the control condition. First, the difference of the subtopic recall rates between the detail cue and control conditions indicate that the participants linked the subtopics with the supporting details in their text memory. This result is consistent with the findings of previous studies that EFL readers integrated supporting details into a subtopic summarizing a paragraph in post-reading tasks (Kim, 2001; Ushiro et al., 2008). Because the subtopics were written at the beginning of each paragraph, the participants might have used them as contexts to understand the incoming supporting details, as theoretical frameworks (e.g., Britton, 1994; Gernsbacher, 1990; Lorch & Lorch, 1985) hypothesis. Moreover, when they finished reading a paragraph and encountered a new subtopic, it is possible that they integrated subordinate propositions (i.e., the supporting details) into superordinate propositions (i.e., the subtopics) in the previous paragraph (e.g., van Dijk & Kintsch, 1983), made a new slot for the new subtopic in their evolving representations of the texts, and located the new subtopic there (e.g., Britton, 1994;

Gernsbacher, 1990).

Additionally, the subtopic recall rates were significantly higher in the major cue condition than in the control condition. This indicates that the participants linked the subtopics with the major topics in their text memory, addressing RQ1 (Can Japanese EFL readers represent topic structure in their text memory?). This result can be explained in a similar way to the links between the subtopics and supporting details, as discussed in the previous paragraph. Britton (1994) assumed that the difference stemmed simply from the size of units in texts. That is, the participants might have regarded the major topics as contexts to understand the following subtopics and then integrated them into the major topics in their evolving representations of the texts (Britton, 1994). Additionally, when finishing reading entire texts, it is possible that they summarized the subtopics into the major topics (Kintsch, 1998).

As shown above, the present result demonstrates readers' memory of links between the major topics and subtopics, which is consistent with findings of previous L1 research (e.g., Hyönä & Lorch, 2004; Lorch & Lorch, 1985, 1996; Lorch et al., 2001; Ritchey et al., 2008; Surber & Schroeder, 2007). These L1 studies found that readers linked paragraphs with the major topics in their text memory but did not directly examine the hierarchical relations between the major topics and the subtopics summarizing these paragraphs. More specifically, these studies determined that a paragraph was recalled when just a single piece of text information was recalled regardless of importance. Thus, the results of the present study add to the findings of previous studies by revealing the direct and specific links between major topics and subtopics.

On the other hand, the results of the present study were inconsistent with those of previous L2 research. Similar to this study, Ushiro et al. (2007) presented important text information as a recall cue but found that it failed to increase participants' recall of texts.

Ushiro et al. (2007) attributed the null effect of their cues to cue redundancy. That is, their participants might have constructed well-structured mental representations, which enabled them to access the cue information even without the recall cues (Rawson & Kintsch, 2004). Additionally, the null effect of their cues might have also occurred because the previous studies did not focus on the links between major topics and subtopics. As some reading models assume, text information is related in accordance with the text's hierarchical structure (e.g., Britton, 1994; van Dijk & Kintsch, 1983). However, because previous L2 studies did not focus on such hierarchical links, the cues might not have had sufficient content overlaps with the recalled information.

These possibilities were also supported by comparing the results of the total recall rates and subtopic recall rates in the immediate recall test of the present study. Whereas the major cues increased subtopic recall, compared with the control condition, the cues did not affect the total recall rates. Like the findings of Ushiro et al. (2007), the effect of the major cues did not appear when the focus was not on the specific links between the major topics and subtopics. Moreover, the participants established well-structured representations of the texts and the major cues might have been redundant (Rawson & Kintsch, 2004). Because of these phenomena, the unfocused content overlaps might have been overridden by the well-structured representations. On the other hand, it should also be noted that the major cues did improve the total recall rates in the delayed recall. Because the participants' text memory faded between the immediate and delayed session, the major cues were not redundant in the delayed recall (Ushiro et al., 2014).

Interestingly, in addition to the fact that the major and detail cues were both effective for the subtopic recall rates, the subtopic recall rates were significantly higher in the major cue condition than in the detail cue condition. This indicated that the participants linked the subtopics to the major topics more strongly than to the supporting details. One reason might be that the participants regarded the links beyond a paragraph to be more important than the links within a paragraph. As the importance rating task confirmed, they identified the different importance of the major topics, subtopics, and supporting details in accordance with topic structure. Another reason might be that the major topics played a more central role in the understanding of the texts. Specifically, whereas the supporting details were just related to the subtopics of the same paragraph, the major topics were directly related to the subtopics and indirectly related to the supporting details via their subtopics, as reflected in the hierarchy of topic structure or the macrostructure of the texts (e.g., van Dijk & Kintsch, 1983). Hence, the major topics were more likely to be represented in text memory than the supporting details, which led the participants to construct stronger links between the major topics and subtopics.

As for when the linking process occurred, there are two possibilities. One possibility is that topic structure processing occurred during reading. Previous research has revealed that L1 readers understand topic structure when they are reading texts (Hyönä & Lorch, 2004; Murray & McGlone, 1997). However, this possibility does not seem valid in Experiment 1. Hyönä et al. (2002) and Hyönä and Nurminen (2006) reported that topic structure processing during reading was not observed among L1 readers with relatively fewer cognitive resources. This might be applied to EFL readers, who need to devote many cognitive resources to lower-level processes (e.g., word recognition and syntactic parsing) and cannot afford to allocate them to higher-level processes including building globally coherent comprehension. In fact, previous research has demonstrated that EFL readers have difficulty linking distant sentences during reading (e.g., Kimura, 2014, 2015b; Morishima, 2013; Ushiro, Nahatame, et al., 2016; Experiment 2) although the focuses of these studies were not topic structure processing. Because many cognitive resources are spent on lower-level processes during L2 reading (e.g., Morishima, 2013), higher-level processes including topic structure processing might be difficult for EFL readers.

Another possibility is that the participants linked major topics with subtopics through answering the recall task after reading. As mentioned in the previous paragraph, cognitive resources available for higher-level processes during EFL reading are limited. However, it is possible that EFL readers can allocate more cognitive resources to higher-level processes in a post-reading task, compared with during reading. This is because resource-demanding processes at word and sentence levels are finished at that time, which makes EFL readers more engaged in higher-level processes (Hosoda, 2014; Nahatame, 2013). For example, although maintenance of global coherence was difficult during EFL reading in Ushiro, Nahatame, et al. (2016; Experiment 2), EFL readers could do so in the recall task after reading in Ushiro, Mori, et al. (2016). They suggested that their participants linked distant sentences during the post-reading task, reconstructing or reorganizing what they had understood in their text memory. Combining the difficulty linking distant text information during EFL reading and reconstructing text memory after reading, the participants of the present study might have been more likely to represent topic structure during the recall task, compared to during reading. However, because the results of the present study alone are not sufficient to determine which possibility is more valid, it will be investigated whether Japanese EFL readers understand topic structure during reading in Experiments 2A and 2B.

Furthermore, L2 reading proficiency did not interact with the effect of the major cues on the subtopic recall rates. This means that the participants were able to link the major topics with the subtopics, regardless of their L2 reading proficiency. Reading proficiency affected understanding of important topics in some studies (Johns & Mayes, 1990; Kimura, 2014) but not in other studies (Ushiro et al., 2009). One possible reason is that reading proficiency likely affects text comprehension more than understanding of important topics, as discussed in Ushiro et al. (2009). In this study, the texts were revised in terms of vocabulary and readability so that the Japanese EFL readers would not have difficulty with literal comprehension. Although the present study examined not only the understanding of important topics but also the links among them, the null effect of L2 reading proficiency might be explained similarly. Another possible reason is that the major cues helped the higher- and lower-proficiency groups establish globally coherent comprehension. In previous studies (Johns & Mayes, 1990; Kimura, 2014), which found a proficiency effect on globally coherent comprehension, the participants were not provided with cues as in the present study. On the other hand, the recall cues were presented in Experiment 1 and then enhanced their memory of topic structure, which might have caused the null effect of proficiency.

3.1.5 Conclusion of Experiment 1

The purpose of Experiment 1 was to examine whether Japanese EFL readers could represent the links between major topics and subtopics in their text memory. Additionally, the participants' memory of links between the subtopics and supporting details was also examined to compare the results between the links beyond and within a paragraph. Furthermore, the total recall rates as well as the subtopic recall rates were calculated to discuss the results in reference to previous studies (e.g., Ushiro et al., 2007) that did not focus on the specific links in accordance with topic structure.

First, the results of the subtopic recall rates indicate that the participants linked the subtopics with the major topics as well as the supporting details in their text memory (RQ1). It is possible that they used the preceding major topics as contexts to understand the following subtopics. They possibly understood the subtopics based on the major topics and summarized the former based on the latter. Additionally, the participants represented more robustly the links between the major topics and subtopics, compared to the links between the subtopics and supporting details. This might have been because major topics are more important information than supporting details. Moreover, more central information, such as major topics, is linked with more text information, which may have caused memory of the major topics to be more robust and available.

Second, the links between the major topics and subtopics might have been represented during the reading or post-reading tasks. Because lower-level processes are not sufficiently developed in most L2 readers, compared with L1 readers, many cognitive resources are consumed in these processes. Hence, L2 readers have difficulty allocating sufficient cognitive resources to higher-level processes including linking text information (e.g., Morishima, 2013). On the other hand, L2 readers can allocate relatively more cognitive resources to higher-level processes after reading because they have finished the resource-demanding lower-level processes. Although Japanese EFL readers are more likely to represent topic structure in their mental representations in a post-reading task, this could not be concluded from the results of Experiment 1 alone. Therefore, the next chapter will examine topic structure processing during EFL reading, adopting an on-line measurement.

3.2 Experiment 2A: On-Line Processing of Topic Structure

3.2.1 Purpose and Research Question

The purpose of Experiment 2A is to examine whether Japanese EFL readers understand the topic structure of expository texts during reading. The results of Experiment 1 suggest that Japanese EFL readers can represent topic structure in their text memory. However, it is still unclear whether the participants understood the topic structure during reading or a post-reading task. Past studies have demonstrated that L1 readers understand topic structure during reading (e.g., Hyönä & Lorch, 2004; Murray & McGlone, 1997). On the other hand, L2 research has not sufficiently investigated whether L2 readers can link subtopics with major topics during reading. Some studies have suggested that L2 readers have difficulty linking distant pieces of text information during reading (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016, Experiment 2) and summarizing text information into an overall theme during reading (Kimura, 2014, 2015b). Combining the findings of previous studies with the results of Experiment 1, it is possible that the participants represented topic structure in their text memory during the cued recall task by reconstructing their text memory rather than linking the subtopics with the major topics during reading. To examine this possibility, Experiment 2 explored topic structure processing during EFL reading. Additionally, the results were compared with during-reading processes to link the subtopics with the supporting details, similar to Experiment 1.

To examine topic structure processing, the priming paradigm (e.g., Meyer & Schvaneveldt, 1971) was adopted in Experiment 2A. The priming paradigm is used to investigate the cognitive processes involved in responses to stimuli, by presenting a priming stimulus and requiring participants to respond to a target probe in a specific way. Although a strict definition of on-line measures does not include the priming paradigm because data is collected in a post-reading task, the present study regarded response times as an on-line measure, following the study of Jiang (2012). To elicit responses to target probes, Experiment 2 adopted the recognition task because respondents need to refer to their text comprehension to make responses and the task was assumed to reflect the priming effect (McKoon & Ratcliff, 1984).

The contents of priming stimuli and target words were determined by adapting the design of Lorch (1993), who examined the integration of text information by interrupting reading halfway and presenting the subtopics of paragraphs, detailed sentences, or no information (i.e., control condition) immediately before resuming reading. Because this study specifically examined whether readers linked subtopics with major topics and supporting details, it presented the major topics (i.e., *major stimuli condition*), supporting details (i.e., *detail stimuli condition*), or information that was topically related to the texts but not suggested explicitly or implicitly in them (i.e., *control stimuli condition*) immediately before the participants recognized the target words. In the control condition, the present study presented information topically related but not suggested explicitly or implicitly in the texts, rather than presenting no information, because the participants would feel strange if priming stimuli preceded only some target words. The control stimuli were expected to be neutral, in contrast to the major and detail stimuli. The results of the recognition task were interpreted based on the following logic. If the participants linked the subtopics with the major topics and supporting details, the major and detail stimuli should improve the participants' recognition of the target words, leading to shorter correct response times and higher correct response rates, compared to the control stimuli. The RQ in Experiment 2 was thus presented as follows:

RQ 2-1: Can Japanese EFL readers process topic structure during reading that is measured by a recognition task?

3.2.2 Method

3.2.2.1 Participants

Fifty-three undergraduate students and graduate students participated in Experiment 2A. Their majors were varied as follows: humanities and culture, human sciences, informatics, life and environmental sciences, medicine and medical sciences, pure and applied sciences, science and engineering, social and international studies, and systems and information engineering. All the participants were native speakers of Japanese who had received at least 6 years of EFL education as part of their formal education in Japan. None of them participated in the other experiments of the present study. Few of them reported their scores on the standardized tests.

3.2.2.2 Materials

The same English-reading proficiency test used in Experiment 1 was adopted in Experiment 2 to measure the participants' discourse-comprehension proficiency. The same experimental texts representing topic structure were used, each including the major topic of the whole text, three subtopics of the paragraphs, and other supporting details. Two texts were paired as a text set, as in Experiment 1. Moreover, a practice text with similar topic structure, text length, and readability was added in Experiment 2A. The practice text was used to familiarize the participants with the experimental procedure on the computer screen. In addition to the practice text, yes–no comprehension questions for the experimental texts, priming sentences, and target and filler words for the recognition task were created in this experiment. These materials will be explained in detail in the following sections.

Yes–no comprehension questions

To motivate the participants to read the texts for comprehension and to assure that they did not have difficulty with literal comprehension, a yes–no comprehension question was created for each paragraph ($k = 3$ for each text). Each question asked about the literal meaning of a sentence representing a supporting detail (see Appendix 2). The participants were told to answer the questions based on what was written in the texts, not based on their own knowledge. The three questions written for each text required both yes and no answers. Two questions required *yes* answers and one question required a *no* answer in half of the texts ($k = 3$), whereas the number of correct *yes* and *no* answers was reversed in the other half of the texts. In total, half of the questions required *yes* answers ($k = 9$) and the other half required *no* answers.

Priming stimuli for the recognition task

To examine whether the participants linked the subtopics with the major topics and supporting details, priming effects between priming stimuli and target words were investigated in Experiment 2A. To this end, three kinds of priming stimuli were created (see Table 3.14): (a) sentences representing the major topics (i.e., *major stimuli*), (b) sentences representing the supporting details (i.e., *detail stimuli*), and (c) sentences that were related to the text topics but not explicitly or implicitly suggested in the texts (i.e., *control stimul*). The detail stimuli represented the supporting details that were not contained in the yes–no comprehension questions. The control stimuli were created to provide a neutral baseline because they were not related to the target words (i.e., the subtopics).

Table 3.14

Priming Stimuli for the Recognition Task in Experiment 2A

Text	Stimuli	*k*	Content
Argentina	Major	1	アルゼンチンには特徴がある
	Detail	3	農場で牛を世話する，製品を外国から輸入する，新聞を自由に出版する
	Control	2	カウボーイが拳銃で決闘する，南米では稲作が盛んだ
Distance	Major	1	距離には要素がある
	Detail	3	エネルギーは移動と関係している，北に十キロ進む行きと帰りで異なる
	Control	2	坂を上るのは大変だ，人混みを割けて迂回する
Energy	Major	1	石油による発電は問題だ
	Detail	3	石油の残りが少ない，売り手の立場が強まる，海の生態系を傷つける
	Control	2	排気に税をかける，エネルギーの自給率を高める
Environment	Major	1	規則で環境を守る
	Detail	3	人々が健康を失う，農薬の流出を制限する，市のサービスで収集する
	Control	2	家の地盤が沈下する，騒音が睡眠を妨げる
Support	Major	1	途上国への支援は問題だ
	Detail	3	お金を条件付きで与える，高価なパーツが必要だ，都市が魅力的に見える
	Control	2	子どもに文字を教える，世界から飢餓をなくす
Three Mile island	Major	1	原発の事故が影響する
	Detail	3	委員会が基準を設ける，新たな安全対策に取り組む，システムの変更が必要だ
	Control	2	世界に技術力をアピールする，壁で放射線を防ぐ

One major stimulus was made for each text (*k* = 6 in the six texts), one detail stimulus for each paragraph (*k* = 18 in total [three paragraphs × six texts]), and two control stimuli for each text (*k* = 12 in total). All the priming stimuli were written in Japanese (i.e., the participants' L1) to avoid the possible effect of surface memory of English expressions. Moreover, all the priming stimuli were written in the present tense and active voice to avoid the possible effect of participants' grammatical knowledge. Further, the priming stimuli were controlled in terms of the amount of information, each containing three content words (i.e., verbs, adjectives, adverbs, or nouns). To confirm whether the priming stimuli represented information as intended, two graduate students who majored in English education classified the priming stimuli into the three categories. The inter-rater agreement rate as the reliability was 91.67%. All the disagreements were resolved in discussion.

Target words and filler words

To measure the activation of the subtopics in the participants' minds, target words representing the central concepts of the subtopics were created. For example, in the case of the subtopic in the distance text *Economic distance is changed by the cost of movement from one place to another*, the central concept was *cost*. The target words were created from each subtopic (*k* = 18 in total [three paragraphs × six texts]). To avoid the possible effect of surface memory, all the target words were translated into Japanese.

The target words were controlled in terms of vocabulary features (i.e., word class, familiarity, frequency, and number of morae; see Table 3.15) because these features might influence the latency and accuracy of responses (de Groot et al., 2002). Regarding word class, all the target words were nouns. Moreover, the target words were high-familiarity words (*M* = 6.21, *SD* = 0.39 in the familiarity ratings using a 7-point Likert scale collected from 40 Japanese university students; Amano & Kondo, 1999). All the target words were also high-frequency words (M = 9.88, SD = 0.95) in the frequency in the newspaper corpus normalized by logarithmic transformation (Amano & Kondo 2000a, 2000b). Word length was controlled from two to four moras (M = 3.11, SD = 0.74).

Table 3.15

Vocabulary Features of Target Words for the Recognition Task

Target word	Text	Mora	Familiarity	Frequency
とち (land)	Argentina	2	6.44	10.67
しげん (resources)		3	6.16	9.04
はってん (development)		4	6.25	10.05
コスト (cost)	Distance	3	6.13	9.48
じかん (time)		3	6.66	10.83
かんかく (feelings)		4	6.16	9.29
きけん (danger)	Energy	3	6.28	9.56
かかく (price)		3	6.19	11.08
かんきょう (environment)		4	6.09	11.57
くうき (air)	Environment	3	6.66	9.55
みず (water)		2	6.53	10.28
ごみ (wastes)		2	6.59	8.93
へいがい (negative effect)	Support	4	5.03	7.99
いぞん (independence)		3	5.50	8.93
じんこう (population)		4	6.53	10.05
いしき (consciousness)	Three Mile island	3	6.25	10.31
けいざい (economy)		4	6.28	11.78
ミス (error)		2	6.16	8.92

To balance the number of *yes* and *no* responses, the same number of filler words that required *no* responses were also created (k = 18). Specifically, the filler words were Japanese words that were related to the text topics but not suggested explicitly or implicitly in the texts. These words were collected as fillers to prevent the participants from making plausibility judgments. If the filler words had been completely unrelated to the text topics, the participants might have made *no* responses to the filler words just because they were not plausible, not because they correctly rejected the words (Hanberlandt, 1994).

Pairs of the priming stimuli and the target/filler words

By pairing the priming stimuli with the target words and filler words, three critical pairs and two filler pairs were created, as shown in Table 3.16. As for the critical pairs, the target words were paired with the major stimuli, detail stimuli, or control stimuli (i.e., *major-sub pairs*, *detail-sub pairs*, and *control-sub pairs*, respectively). These critical pairs were made for each paragraph of the six texts (k = 18 in total [three paragraphs × six texts]).

Table 3.16

Sample of the Critical and Filler Pairs for the Recognition Task

Pair	*k*	Priming stimuli	Target/filler word	Correct response
Major-sub	18	アルゼンチンには特徴がある	しげん	Yes
Detail-sub	18	農場で牛を世話する	とち	Yes
Detail-filler	12	製品を海外から輸入する	いみん	No
Control-sub	18	南米では稲作が盛んだ	はってん	Yes
Control-filler	6	カウボーイが拳銃で決闘する	おんだん	No

The major-sub pairs were assumed to extract activation of links between the major topics and subtopics, whereas the detail-sub pairs were to extract activation of links between the supporting details and subtopics. The control-sub pairs were designed as a baseline condition for the major and detail pairs. By comparing the latency and accuracy of responses to the major-sub and detail-sub pairs with those of the control-sub pairs, it was analyzed whether the participants linked the subtopics with the major topics and supporting details.

Two kinds of filler pairs that required *no* responses were also created. In the detail-filler pairs, the detail stimuli were paired with the filler words. In other words, the detail-filler pairs were composed of the filler words and priming stimuli that were mentioned in the texts. Without the detail-filler pairs, the participants might think that the correct responses were always *yes* when the priming stimuli consisted of information explicitly stated in the texts. In the control-filler pairs, the control stimuli were paired with the filler words. Without the control-filler pairs, the participants might expect that the correct responses were always *yes* when the priming stimuli consisted of information that was not suggested in the texts. To balance the number of correct *yes* and *no* responses, two detail-filler pairs and one control-filler pair were created for each text. This resulted in a total of 12 detail-filler pairs and 6 control-filler pairs for the 6 texts.

Three sets of the critical and filler pairs were created for counterbalance, using the Latin-square design. Each of the priming-pair sets contained 18 critical pairs (i.e., 6 major-sub pairs, 6 detail-sub pairs, and 6 control-sub pairs) and 18 filler pairs (i.e., 12 detail-filler pairs and 6 control-filler pairs). Three types of critical pairs were assigned to three paragraphs in each text. For example, in the case of Priming-Pair Set 1 for the two texts of Text Set A, the detail-sub pairs were made from the first paragraphs, major-sub pairs from the second paragraphs, and control-sub pairs from the third paragraphs. Assignment of the critical pairs to paragraphs 1-3 was cycled across Text Sets A, B, and C, as illustrated in Table 3.17. Assignment of each paragraph to critical pairs was cycled across the three priming-pair sets.

Table 3.17

Assignment of the Critical Pairs to Each Paragraph in Priming-Pair Set 1

	Paragraph 1	Paragraph 2	Paragraph 3
Text Set A	Detail-sub pair	Major-sub pair	Control-sub pair
Text Set B	Control-sub pair	Detail-sub pair	Major-sub pair
Text Set C	Major-sub pair	Control-sub pair	Detail-sub pair

Pilot study

A pilot study was conducted to confirm whether the critical pairs themselves varied in terms of the degree of semantic priming between the priming stimuli and the target words (i.e., to what degree the priming stimuli and target words were semantically related) when the participants did not read the texts. Specifically, it was verified if the correct response times and rates to the target words were similar among the three kinds of critical pairs. By doing so, it would be possible to interpret the results in the main experiment as follows. Differences in the correct response times and rates among the critical pairs would indicate successfully represented links that connected the subtopics with the major topics and supporting details, and not simply the different semantic priming that the critical pairs themselves originally had.

In the pilot study, 15 Japanese EFL university students were tested individually. Because the participants did not read the texts, a lexical decision task was conducted instead of the recognition task. Hence, the filler words were changed into pseudo words by replacing one letter with another in a manner that preserved their pronounceability (Jiang, 2012). The pilot experiment was administered on a computer using SuperLab 5.0 software (Cedrus, US). One of the three priming-pair sets was randomly assigned to each of the pilot participants.

The task began with the word “Ready?” displayed in the center of the computer screen. The pilot participants pushed the *yes* button on the response pad (RB-740 model, Cedrus, US) to start the test. A fixation point (***) was presented in the center of the screen for 500 milliseconds (ms), followed by a priming stimulus for 1,250 ms. These amounts of time were determined during the pre-testing to ensure that the participants had sufficient time to recognize the priming stimuli. After the priming sentence, a blank screen appeared for 300 ms, followed by either a target word or filler word, which remained on the screen until the participant responded. The participants pushed the *yes* or *no* button as quickly and accurately as possible to indicate their lexical decision, and their correct response times and rates were recorded. The trials described above were repeated for the 36 critical pairs in random order. Before the experimental session, the pilot participants had a practice session.

The results showed that the correct response rates were quite high for the critical pairs; the participants responded correctly to 98.89% of them (89 out of 90 data points). Table 3.18 illustrates descriptive statistics of the correct response times. The correct response times deviating more than 2.5 standard deviations (*SD*s) from the mean of each individual participant were replaced with the mean plus 2.5 *SD*s. This means that the criteria of the outliers differed for each participant. This resulted in the replacement of 2.22% of the data (6 of 270 data points).

Table 3.18

Descriptive Statistics of the Correct Response Times (ms) of the Lexical Decision Task in the Pilot Study of Experiment 2A

Critical pair	*M*	95% CI	*SD*
Major-sub	632	[569, 694]	113
Detail-sub	622	[560, 683]	110
Control-sub	634	[569, 699]	117

A one-way repeated ANOVA was conducted for the correct response times with the critical pairs (i.e., major-sub, detail-sub, control-sub) as a within-participant factor. This analysis did not find any differences among the three critical pairs, $F(2, 28) = 0.30$, $p = .742$, $\eta_p^2 = .02$. This confirmed that semantic priming did not differ among the experimental and control pairs.

3.2.2.3 Procedure

The participants were tested individually. First, the experimenter gave oral and written explanations of the research purpose and the procedure and obtained informed consent from the participants. The participants then filled out the profile sheet including a self-report of scores on large standardized tests. The main experiment was a single session that lasted for about 70 minutes and included the reading session, the importance rating task, and the English-reading proficiency test. Figure 3.5 illustrates the procedure of the reading session.

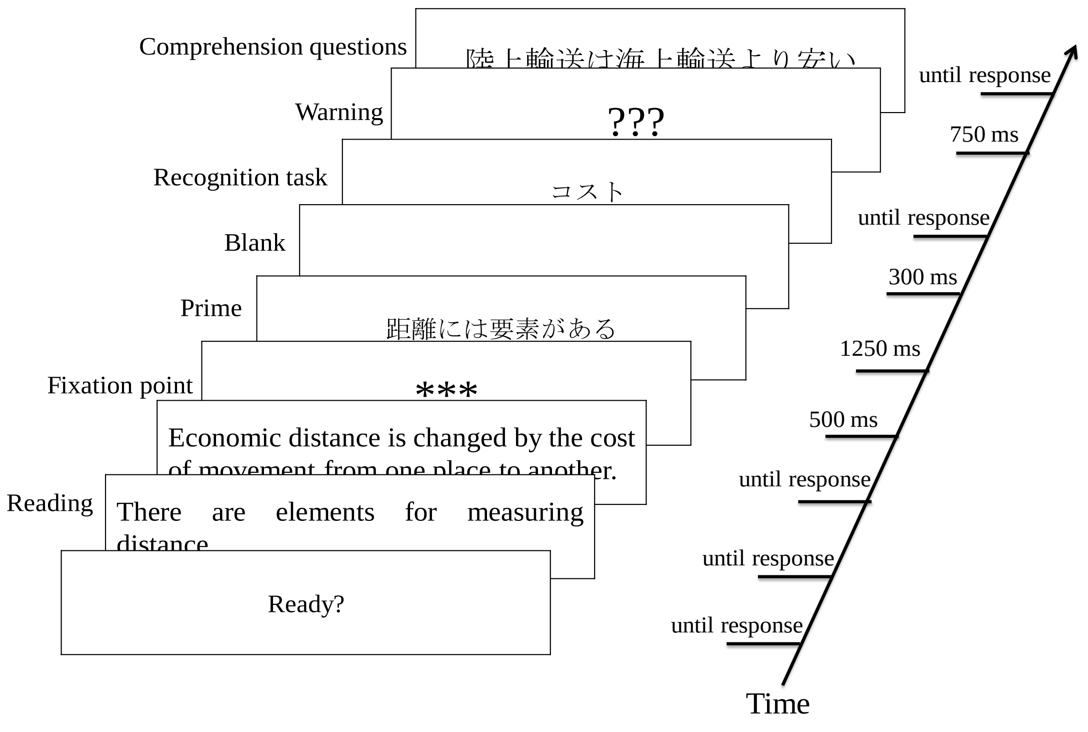

Figure 3.5. Procedure of the reading session in Experiment 2A.

In the reading session, the participants were instructed to read each text for comprehension to answer the yes–no comprehension questions. They read the text on the screen sentence by sentence at their own pace, pressing the *yes* button to proceed to each following sentence. After reading each text, a fixation point, a priming sentence, a blank screen, and a target or filler word were presented as in the pilot study. The participants pushed the *yes* or *no* button to determine whether the target or filler word was explicitly written in the text (i.e., the recognition task). This sequence was repeated for six trials (i.e., one major pair, one detail pair, one control pair, one filler 2 pair, and two filler 1 pairs). The presentation order of the pair types for each text was counterbalanced across the pair sets. After the recognition task, a warning for the yes–no comprehension questions ("???") was presented for 700 ms. Three yes–no comprehension questions then appeared onscreen one at a time. After the participants answered these questions by pushing the *yes* or *no* button, the accuracy feedback was presented for 2,000 ms to motivate the participants to read the texts for comprehension. The above procedure was repeated for each of the six texts, with the text order randomized. A practice session was conducted before the main session to familiarize the participants with the procedure.

After the reading session, the participants rated the importance of each sentence using 5-point Likert scale. As in Experiment 1, the importance ratings were collected to confirm that the participants correctly identified the topic structure as intended. After finishing the importance rating, the participants had a 5-minute break and then began the English-reading proficiency test within 30 minutes.

3.2.2.4 Scoring

Prior to calculating the average correct response latency, the correct response times deviating more than 2.5 *SD*s from the mean of an individual participant were replaced with the response time deviating 2.5 *SD*s from the mean. This resulted in the replacement of 2.31% of the data (18 out of 780 responses). The correct response times and rates were averaged for each condition across the proficiency groups and the three critical pairs. In Experiment 2A, the correct response times were not divided by the length of the target words (e.g., the number of morae) based on Trueswell, Tanenhaus, and Garnsey (1994). They found that the reading times did not change linearly with the number of characters; specifically, the reading times per character were longer for smaller information units. Although their finding was not about response times, a similar phenomenon seemed to occur because the respondents first needed to read the target words before making responses.

3.2.2.5 Analysis

To confirm the validity of the topic structures of the texts, a 2 (proficiency: higher, lower) × 3 (information: major topic, subtopic, supporting detail) two-way mixed ANOVA was carried out for the importance ratings, with proficiency as a between-participant factor, and information as a within-participant factor. Furthermore, addressing the RQ, 2 (proficiency: higher, lower) × 3 (pair: major-sub, detail-sub, control-sub) two-way mixed ANOVAs were performed for the correct response times and rates, with proficiency as a between-participant factor, and pair as a within-participant factor.

3.2.3 Results

3.2.3.1 English-Reading Proficiency Test

Table 3.19 illustrates the descriptive statistics of the English-reading proficiency test (Cronbach's α = .820). The participants were divided into a higher-proficiency group (n = 29) and a lower-proficiency group (n = 24) based on the median of their scores. A t test yielded a significant difference for the scores of the two proficiency groups, $t(51) = 9.91$, $p < .001$, d =

2.74. This result confirmed that the higher-proficiency group outperformed the lower-proficiency group on the English-reading proficiency test.

Table 3.19

Descriptive Statistics of the English-Reading Proficiency Test in Experiment 2A

		Second (k = 20)		Pre-first (k = 8)		Total (k = 28)	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	29	16.24	1.90	2.38	2.06	18.62	2.82
Lower	24	10.42	2.80	0.50	0.78	10.92	2.81

Similar to Experiment 1, the English-reading proficiency of the higher-proficiency group was estimated to be between the the second and pre-first grades of the EIKEN test (levels B1 and B2 on the CEFR; Dunlea, n.d.). Their correct answer rates were higher than 60% for the test items of the second grade but lower than 70% for pre-first grade test. On the other hand, the English-reading proficiency of the lower-proficiency group was estimated to be between the pre-second and second grade (levels A2 and B1 on the CEFR; Dunlea, n.d.). Their correct answer rates were lower than 70% for the second grade test and close to 60% for the pre-first grade test.

3.2.3.2 Importance Ratings

To confirm that the participants identified the topic structures of the texts as intended, an importance rating task was conducted. Table 3.20 and Figure 3.6 show the descriptive statistics of the importance ratings. A 2 (proficiency: higher, lower) × 3 (information: major topic, subtopic, supporting detail) two-way mixed ANOVA was performed for the importance ratings, with proficiency as a between-participant factor and information as a within-participant factor. As a result, the interaction was not significant, $F(1.62, 82.50) = 1.69$, $p = .195$, $\eta_p^2 = .03$, nor was the main effect of proficiency, $F(1, 51) = 0.32$, $p = .573$, $\eta_p^2 = .01$.

Table 3.20

Descriptive Statistics of the Importance Ratings in Experiment 2A

		Major topic		Subtopic		Detail	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	29	4.26	0.66	3.75	0.34	2.86	0.38
Lower	24	4.04	0.54	3.73	0.36	2.97	0.35

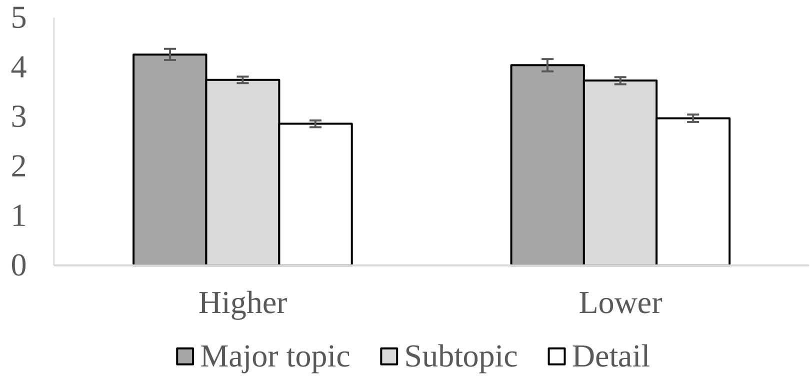

Figure 3.6. Mean importance ratings in Experiment 2A (± *SEM* bars).

However, the main effect of information was significant, $F(1.62, 82.50) = 97.22$, $p < .001$, $\eta_p^2 = .66$. Multiple comparisons revealed that the participants judged that the major topics were significantly more important than the subtopics, $p < .001$, $M_{\text{diff}} = 0.41$, 95% CI [0.19, 0.63]; and that the subtopics were significantly more important than the supporting details, $p < .001$, $M_{\text{diff}} = 0.82$, 95% CI [0.65, 1.00]. These results confirmed that the participants identified the major topics, subtopics, and supporting details as the raters intended. The results of the ANOVA are summarized in Table 3.21.

Table 3.21

Summary of the Two-Way ANOVA for the Importance Ratings in Experiment 2A

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.06	1.00	0.06	0.32	.573	.01
Error (P)	10.08	51.00	0.20			
			Within-participants			
Information (I)	41.60	1.62	25.71	97.22	< .001	.66
I × P	0.72	1.62	0.45	1.69	.195	.03
Error (I)	21.82	82.50	0.26			

3.2.3.3 Yes–No Comprehension Questions

Yes–no comprehension questions were conducted to motivate the participants to read the texts for comprehension and to confirm that they did not have difficulty with literal comprehension. Table 3.22 shows the descriptive statistics of the yes–no comprehension questions. The table illustrated that the participants correctly answered more than 80% of the questions, regardless of the proficiency group. This result confirmed that they read for comprehension and sufficiently understood the texts.

Table 3.22

Descriptive Statistics of the Yes–No Comprehension Questions in Experiment 2A

	n	*M*	*SD*	Max	Min	95% CI
Higher	29	.86	.10	1.00	.61	[.82, .90]
Lower	24	.84	.09	1.00	.67	[.80, .87]

3.2.3.4 Correct Response Times of the Recognition Task

To address the RQ, the correct response times were compared among the three critical pairs. A 2 (proficiency: higher, lower) × 3 (pair: major-sub, detail-sub, control-sub) two-way mixed ANOVA was conducted. The results demonstrated that none of the interaction or main effects were significant (all *p*s > .10). This result indicates that both proficiency groups did not make correct responses to the major-sub pairs or the detail-sub pairs more quickly, compared to the control-sub pairs (see Table 3.23 and Figure 3.7). The participants did not strongly activate the links among the major topics, subtopics, and supporting details, regardless of their proficiency. Table 3.24 summarizes the results of the two-way ANOVA.

Table 3.23

Descriptive Statistics of the Correct Response Times (ms) *of the Recognition Task*

		Major-sub		Detail-sub		Control-sub	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	29	1257	387	1262	383	1272	591
Lower	24	1221	433	1216	539	1303	807

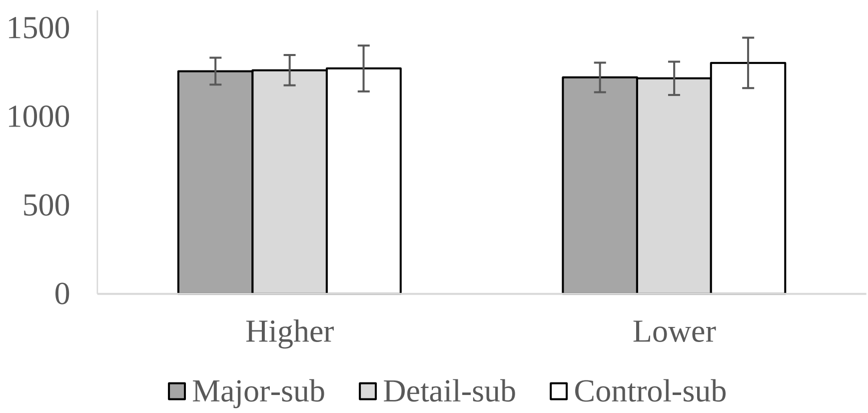

Figure 3.7. Mean correct response times (ms) of the recognition task (± *SEM* bars).

Table 3.24

Summary of the Two-Way ANOVA for the Correct Response Times of the Recognition Task

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-Participants			
Proficiency (P)	10877.18	1.00	10877.18	0.02	.902	< .01
Error (P)	36433539.01	51.00	714383.12			
			Within-Participants			
Pair (P')	82237.58	1.39	59159.11	0.55	.517	.01
P × P'	46537.36	1.39	33477.51	0.31	.653	.01
Error (P')	7637775.38	70.90	107732.82			

3.2.3.5 Correct Response Rates of the Recognition Task

To address the RQ, the correct response rates were also examined. Table 3.25 and Figure 3.8 show the descriptive statistics of the correct response rates. A 2 (proficiency: higher, lower) × 3 (pair: major-sub, detail-sub, control-sub) two-way mixed ANOVA was performed for the accuracy data. As a result, neither the interaction nor the main effect of proficiency was found to be significant (all *p*s > .10).

Table 3.25

Descriptive Statistics of the Correct Response Rates of the Recognition Task

		Major-sub		Detail-sub		Control-sub	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	29	.89	.11	.88	.12	.75	.24
Lower	24	.89	.11	.81	.18	.72	.29

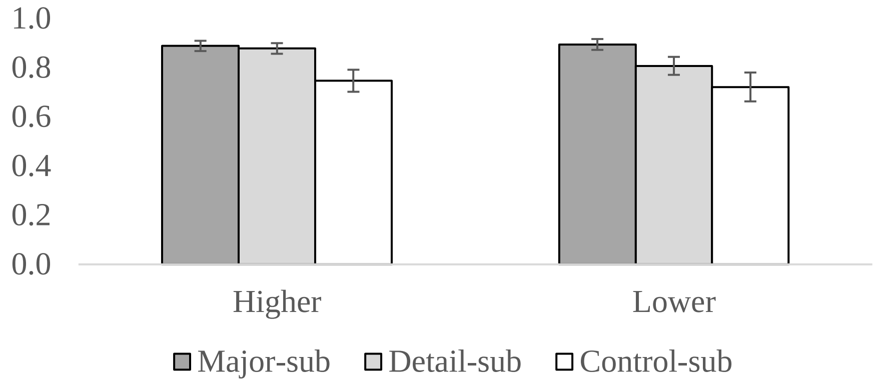

Figure 3.8. Mean correct response rates of the recognition task (± *SEM* bars).

However, a significant main effect of the pairs was yielded, $F(1.60, 81.64) = 9.39$, $p = .001$, $\eta_p^2 = .16$. Multiple comparisons revealed that the participants made significantly more correct responses to the major-sub pairs, $p = .001$, $M_{diff} = 0.16$, 95% CI [0.06, 0.26], and to the detail-sub pairs, $p = .043$, $M_{diff} = 0.11$, 95% CI [0.00, 0.22], compared to the control-sub pairs. However, there was no significant difference between the major-sub pairs and detail-sub pairs, $p = .217$, $M_{diff} = 0.05$, 95% CI [–0.02, 0.11]. Table 3.26 summarizes the ANOVA results.

Table 3.26

Summary of the Two-Way ANOVA for the Correct Response Rates of the Recognition Task

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	0.04	1.00	0.04	1.15	.289	.02
Error (P)	1.57	51.00	0.03			
			Within-participants			
Pair (P’)	0.68	1.60	0.43	9.39	.001	.16
P × P’	0.04	1.60	0.02	0.54	.546	.01
Error (P’)	3.70	81.64	0.05			

3.2.4 Discussion

Three possibilities for the disagreement between the correct response times and rates

The results of Experiment 2A were complicated because the results of the correct response times and rates were not consistent. That is, the correct response times did not differ among the three critical pairs. This indicates that links among the major topics, subtopics, and supporting details were not sufficiently activated in the participants' minds or were made available for the recognition task immediately following reading. On the other hand, the correct recognition rates were significantly higher for the major-sub and detail-sub pairs than for the control-sub pairs, suggesting that the participants recognized links among the major topics, subtopics, and supporting details in their mental representations. These conflicting results regarding recognition speed and accuracy might be explained by the following three possibilities. RQ2-1 will be addressed after a discussion of these possibilities.

The first possibility is that the participants linked the subtopics with the major topics and supporting details during reading, as reported by previous L1 studies (e.g., Hyönä & Lorch, 2004; Murray & McGlone, 1997), but such understandings had been deactivated by the time of the recognition task. This possibility is consistent with Gernsbacher's (1990) structure building hypothesis. This theoretical framework hypothesizes that when readers have finished reading a unit of a text (e.g., a paragraph), they link the information in it to a superordinate topic summarizing it and deactivate comprehension of the previous unit. Because readers have a limited amount of cognitive resources, they must preserve some to process the incoming unit. Although the abovementioned frameworks are based on L1 readers, their assumptions may also be applied to L2 readers. Cognitive resources available in L2 reading are most likely even more strictly constrained than in L1 reading, because L2 readers are not sufficiently proficient in lower-level processes (e.g., word recognition and syntactic parsing), compared to L1 readers (e.g., Morishima, 2013; Yoshida, 2003). Based on the first possibility, the participants of the present study might have linked the subtopics with the major topics and subtopics, but deactivated the links when they encountered a new paragraph about another subtopic and noticed that the previous paragraph had ended.

The second possibility is that the participants linked the subtopics with the major topics and supporting details to some extent, but they were not able to establish links sufficiently. Hence, the links might have not been strongly activated in their minds at the time of the recognition task. Because of this, the links were not immediately available during the recognition task even when the major and detail stimuli were presented, but the participants were able to recognize them within their mental representations. Because the links were not robustly established in the text information, presenting the major and detail stimuli led to more accurate responses to the target words than the control stimuli, but did not yield more accurate and faster responses. This possibility is consistent with that of the fact that L2 readers have limited cognitive resources available for higher-level processes during reading (e.g., Morishima, 2013; Yoshida, 2003).

Concerning the second possibility, the participants' insufficient understanding of the links among text information might have been caused by difficulty understanding individual pieces of text information or difficulty connecting them. However, the data in this study seem to refute this hypothesis. Specifically, the correct answer rates of the yes–no comprehension questions were over 80% for both proficiency groups, and the correct response rates to the target words (i.e., the subtopics) were over 70% regardless of the proficiency group or critical pair. These data suggest that the participants did not have difficulty understanding individual sentences including the subtopics. Thus, in case of the second possibility, despite understanding individual sentences, the participants might have failed to link the subtopics with the major topics and supporting details, leaving these pieces of text information fragmented in their minds.

The third possibility is that different cognitive processes might have been reflected in the correct response times and rates. Jiang (2012) stated that task performance is thought to reflect on-line processes when fast responses are required and response times are measured. Based on this statement, while the correct response times were indicative of during-reading processes, the correct response rates did not necessarily reflect these processes, including cognitive processes induced by the recognition task itself. As a characteristic process induced by the recognition task, the task required the readers to refer to their text memory and judge whether the target probes had been included in it (e.g., McKoon & Ratcliff, 1984). This might have helped the participants to reconstruct their text memory (e.g., Ushiro, Mori, et al., 2016) and then link the major topics, subtopics, and supporting details.

Different effects of the major topics and supporting details on subtopic retrieval in participant's recall and recognition rates

As discussed above, the third possibility assumes the reconstruction of text memory in the recognition task, similar to the recall task in Experiment 1. However, the results of the correct recognition rates in Experiment 2A and the subtopic recall rates in Experiment 1 showed different tendencies. That is, while the participants remembered the subtopics better with the major cues than with the detail cues in Experiment 1, they remembered the subtopics to a similar extent when the major stimuli and detail stimuli were presented in Experiment 2A. Two reasons could explain the difference between the results in the two experiments.

First, different features of the recall and recognition tasks might have caused the different results. Although both the recall and recognition tasks required the participants to remember what they had comprehended in the texts, quick responses were required in the recognition task alone. Hence, the participants in Experiment 2A were not likely to have sufficient time for the reconstruction of text memory. The limited answer time might have allowed the participants to link the subtopics and supporting details within paragraphs but not the subtopics with the major topics beyond paragraphs, which requires the integration of more distant text information to do so sufficiently.

Second, the productiveness of the recall task and the receptiveness of the recognition task might have led to different results between the experiments. This is because the recall task required the participants to reproduce their text comprehension on their own and thus to reorganize it (e.g., Ushiro, Mori, et al., 2016). In contrast, the recognition task was receptive, requiring only yes–no judgments. Hence, it did not seem to require the participants to reorganize or reconstruct their text memory, compared to the productive recall task. This view is consistent with activity theory, which hypothesizes that deep-level text comprehension (e.g., selection and organization of information) can be achieved by completing a productive task (e.g., Stull & Mayer, 2007). To understand topic structure, readers need to identify the major topics and subtopics in texts (corresponding to the selection of information) and then establish hierarchical links between them (corresponding to the organization of information), which were more likely to appear in productive tasks than in receptive tasks.

These possibilities did not seem to conflict in interpreting the different effects of the major topics and supporting details on the subtopic retrieval in the recall and recognition rates. Thus, it is possible that both of the possibilities interacted with each other.

Interpretations of understandings of links among text information based on the three possibilities (RQ2-1)

Although three possibilities were discussed above, different possibilities might have been more valid, depending on the type of link. As for the links between the subtopics and supporting details, the first of the three possibilities seems most valid because it is consistent with past findings. For example, Mori (2015) indicated that the Japanese university students activated the subtopics by mentally summarizing the paragraphs in their minds while reading. Although Mori's (2015) study did not focus directly on links between the subtopics and supporting details, understandings of the subtopics required integrating supporting details into a superordinate proposition (e.g., Britton, 1994; van Dijk & Kintsch, 1983). Thus, the activation of such subtopics suggested that the participants linked the supporting details with the subtopics during reading. Additionally, although the focus was different from this study, Kimura (2013) also demonstrated that Japanese university students understood themes summarizing a paragraph-long narrative. The participants read the last sentences of the narratives faster when they represented the themes than when they did not. This was probably because the participants were developing understandings of the theme during reading and integrating the sentences into the evolving theme comprehension. These previous findings seem to support the idea that the participants of the present study linked the subtopics to the major topics during reading and then deactivated them because of limited cognitive resources (e.g., Gernsbacher, 1990), which explains why these links were not immediately available but were correctly remembered in the recognition task.

As for L2 reading proficiency as a reader factor, this did not interact with recognition times or accuracy. The null effect of L2 reading proficiency was consistent with Experiment 1. Similar to Experiment 1 and previous L2 studies (Ushiro et al., 2009), L2 reading proficiency might affect text comprehension more than the participants' ability to link the subtopics with other text information. This view is supported by the data collected in this study. The correct answer rates of the yes–no comprehension questions and the correct response rates of the recognition task confirmed that the participants comprehended the texts sufficiently. Combined with the findings of previous studies, it can be said that the effect of L2 reading proficiency did not appear because the participants sufficiently understood the texts, regardless of the proficiency group.

On the other hand, the first possibility is less likely than the second and third possibilities to explain the results concerning the links the participants made between the major topics and subtopics. Although previous studies have demonstrated that readers often understand topic structure during reading, the participants were adult L1 readers (i.e., university students) in most of the studies (e.g., Murray & McGlone, 1997). Hyönä and his colleagues examined individual differences in L1 reading processes, indicating that readers with few cognitive resources did not link the subtopics with the major topics during reading (Hyönä et al., 2002; Hyönä & Nurminen, 2006). This is also probably true of EFL readers. Because they are not proficient in lower-level cognitive processes during reading (e.g., word recognition, syntactic parsing), they must spend many cognitive resources on these basic processes during reading (e.g., Morishima, 2013), which compete for cognitive resources available for higher-level processes, such as topic structure processing. Although few L2 studies have examined topic structure processing during reading, some L2 studies have reported that EFL readers failed to link distant sentences during reading (Morishima, 2013; Ushiro, Nahatame, et al., 2016, Experiment 2) and comprehend themes beyond paragraphs in expository texts (Kimura, 2014). Taken together with the results of the present study, the participants might have had difficulty with topic structure processing.

However, during the recognition task after reading, the participants might have had more cognitive resources available for the linking processes. This is because the lower-level cognitive processes had been finished before the post-reading task, as discussed in Experiment 1. Hence, even though higher-level processes were difficult during reading, these processes might have been relatively easy after reading. The second and third possibilities are also consistent with the findings of previous research. For example, Japanese university students failed to achieve coherent comprehension between distant sentences during reading in Ushiro, Nahatame, et al. (2016, Experiment 2), but such readers could do so in the recall task after reading (Ushiro, Mori, et al., 2016). Similarly, inference generation that was difficult during EFL reading was observed in post-reading tasks in some L2 studies (Hosoda, 2014; Nahatame, 2013).

Regarding linking the subtopics with the major topics, the proficiency effect did not appear. However, the reasons for the null effect seemed different from linking the subtopics with the supporting details. That is, the participants might have understood the links within paragraphs (i.e., the first possibility), regardless of L2 reading proficiency. On the other hand, both proficiency groups likely had difficulty comprehending the links across paragraphs (i.e., the second and third possibilities). In Ushiro et al.'s (2009) study on globally coherent comprehension, the participants wrote good summaries, and thus the proficiency effect did not influence their summary performance. In contrast, the participants in Johns and Mayes (1990) were able to produce few macropropositions in the summary task, regardless of L2 reading proficiency. As shown in previous studies, different factors might result in a null proficiency effect.

Furthermore, regarding the third possibility, to determine whether the result of Experiment 2 was the outcome of a specific task or a more general phenomenon of EFL reading, a follow-up study seems necessary. Because the third possibility assumed understandings of topic structure was induced by the recognition task, a follow-up study with a different task might be necessary. Jiang (2012) recommended modified replication adopting different participants, materials, tasks, etc. If the results of the correct response times and rates for the three critical pairs were due to the task-induced process (i.e., reference to and reconstruction of text memory), different tendencies should be observed in different tasks. On the other hand, if the results of Experiment 2 reveal general phenomena of Japanese EFL readers, a similar tendency would be expected to appear. As reviewed in Chapter 2, the lexical decision task seems appropriate for a follow-up study. This task is most often adopted to measure L2 reading comprehension (e.g., Jiang, 2012) and does not induce reference to text memory; rather, it simply reflects what is activated at the time of the task. This will be further explored in the next experiment.

3.2.5 Conclusion of Experiment 2A

The purpose of Experiment 2A was to investigate whether Japanese EFL readers linked the subtopics with the major topics and supporting details during reading. To this end, the correct response times and rates to the target words (i.e., subtopics) were compared when the major topics, supporting details, and topically related but unsuggested information were presented as the priming stimuli. The main findings of this experiment can be summarized into the following two points: (a) Links among text information were not strongly activated or immediately available in the participants' minds during the recognition task, but (b) they remembered links between the major topics and subtopics and links between the subtopics and supporting details. These results can be interpreted in three possible ways.

First, the participants initially built links among text information during reading, but these understandings had been deactivated by the time of the recognition task. When finishing reading a unit in the texts and upon encountering a new unit, it is assumed that the readers deactivated their understanding of what they had understood in the previous unit. This process is hypothesized because readers have a limited amount of cognitive resources, which they must reserve for reading the new unit. This seems to apply to this study because EFL readers' cognitive resources are more limited than L1 readers'.

Second, although the participants did not have difficulty with individual text comprehension, they had difficulty establishing robust links among the text information while reading. Because of this, the participants were likely to only weakly activate these links in their minds. Hence, because the links were not readily available in the recognition task, the major and detail stimuli helped the participants to respond accurately, but not both accurately and quickly.

Third, the participants might have failed to link text information during reading but were induced to engage in the process during the recognition task. The recognition task required them to refer to their text memory to answer the task questions. This might have triggered the reconstruction of text memory, allowing the participants to link text information, which was difficult during reading.

The first possibility is likely to explain the links made by the participants between the subtopics and supporting details. This is because previous studies have suggested that EFL readers seem to have the ability to connect supporting details to subtopics during reading (Kimura, 2013; Mori, 2015). On the other hand, the second and third possibilities were likely to be applied to links between the major topics and subtopics. Whereas EFL readers are strictly limited in cognitive resources during reading, they have relatively more resources in post-reading tasks after finishing the resource-demanding lower-level processes. Hence, they can engage in the higher-level processes (e.g., establishing globally coherent comprehension), which may be difficult during EFL reading, when tackling post-reading tasks by reconstructing their text memory.

However, as suggested by the third possibility, the participants' understandings of topic structure might have been affected by the recognition task. Thus, a follow-up study will be conducted in the next section, adopting the lexical decision task instead of the recognition task. If the third possibility regarding the task-induced process is valid, differences in correct response rates between the major and control pairs would not be observed. On the other hand, if the participants understood topic structure regardless of the priming task, similar results for the correct response rates would be expected to be observed. This will be further explored in Experiment 2B.

3.3 Experiment 2B: Follow-Up Study of Experiment 2A

3.3.1 Purpose and Research Question

The purpose of Experiment 2B was to replicate Experiment 2A and examine whether Japanese EFL readers linked the subtopics with the major topics and supporting details while reading. In Experiment 2A, the correct response times did not differ among the critical pairs, whereas the correct response rates were higher for the major-sub and detail-sub pairs than for the control-sub pairs. These conflicting results were discussed in terms of the following three possibilities: (a) the participants linked text information during reading and deactivated that understanding, (b) the heavy use of cognitive resources during reading led to the insufficient availability of cognitive resources for the linking processes, and (c) the recognition task required the participants to refer to their text memory and then induced its reconstruction, which contributed to linking text information. The first possibility may apply to the readers' understanding of links between the subtopics and major topics. This possibility is consistent with past research concerning EFL readers' ability to link text information within a paragraph during reading (e.g., Kimura, 2013; Mori, 2015). On the other hand, the second and third possibilities may explain the readers' understanding of links between the major topics and subtopics, which is consistent with previous studies that found constrained cognitive resources available for higher-level processes during EFL reading (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016) and relatively more cognitive resources available for higher-level processes in post-reading tasks (e.g., Ushiro, Mori, et al., 2016). Because the recognition task might have affected reader understanding of the topic structure, a replication with a different priming task was conducted in Experiment 2B. If the understanding of topic structure was induced by the recognition task, it should not appear in the lexical decision task. Moreover, if the readers' understanding of topic structure was due to during-reading processes rather than task-induced processes, similar results should be observed in Experiment 2B.

In this experiment, a lexical decision task was adopted as a priming task. In previous research on text comprehension, lexical decision tasks have been the most widely used, in addition to recognition tasks (Jiang, 2012). A lexical decision task asks whether a target probe is a real word, and does not require reference to text memory or other processes related to text comprehension. The process of interpreting the correct response times and rates was the same as that used in Experiment 2A. If the participants linked the subtopics with the major topics and supporting details during reading, the understandings of these links should have been activated in their minds. This activation should enable the major and detail stimuli to improve lexical decision performance for the target words. Hence, correct response times should be shorter and correct response rates should be higher than for the control stimuli. Therefore, the RQ for this experiment was as follows:

RQ2-2: Can Japanese EFL readers process topic structure during reading that is measured by a lexical decision task?

3.3.2 Method

3.3.2.1 Participants

Thirty-four Japanese undergraduate students and graduate students participated in the experiment. Their majors were varied, as follows: human sciences, humanities and culture, informatics, life and environmental sciences, medicine and medical sciences, science and engineering, and social and international studies. All the participants were native speakers of Japanese who had received at least six years of EFL education as part of their formal education in Japan. None of them participated in any of the other experiments in the present thesis. Based on CEFR alignment studies (Dunlea, n.d.; ETS, 2015, 2017), their general English proficiency was estimated to range from the beginner to the advanced level (i.e., about A1 to C1 levels on the CEFR), according to self-reported scores for the TOEIC listening and reading test (range: 520–935; 501–600: n = 1; 601–700: n = 2; 701–800: n = 1; 901–990: n = 1), the TOEFL ITP test (range: 440–550; 401–500: n = 4; 501–600: n = 5), and the EIKEN grades (range: 3rd–1st; 3rd: n = 3; pre-2nd: n = 5; 2nd: n = 9; pre-1st: n = 2; 1st: n = 1). It should be noted that some participants did not report any of their TOEIC, TOEFL ITP, or EIKEN scores, whereas other participants reported all the above test scores.

3.3.2.2 Materials

To replicate Experiment 2A, the following materials were adopted from that experiment (see Appendix 2): (a) the experimental and practice texts including topic structure; (b) the yes–no comprehension questions querying literal understandings of supporting details in the texts; (c) the priming stimuli representing the major topics, supporting details, and topically related but unsuggested information; (d) the target words representing the central concepts of the subtopics; and (e) the English-reading proficiency test to measure discourse-comprehension skills. However, one change was made in Experiment 2B. Because the recognition task was replaced with the lexical decision task in this experiment, the filler words (i.e., the information that was related to the text topics but not explicitly or implicitly suggested in the texts) were transformed into pseudo words. Specifically, one letter of each filler word was substituted for another letter to keep the filler words pronounceable (Jiang, 2012), as in the pilot study of Experiment 2A. The pseudo words were paired with the detail stimuli and the control stimuli to create the filler pairs, requiring *no* responses in the lexical decision task.

3.3.2.3 Procedure

The procedure of Experiment 2B was the same as that of Experiment 2A, apart from the two following changes: (a) The participants answered the lexical decision task after reading each text, and (b) they did not rate the importance of each sentence after reading all the texts. In the experiment, the participants individually read the texts for comprehension sentence by sentence at their own pace on a computer screen. After reading each text, a major stimulus, detail stimulus, or control stimulus was presented and the participants had to determine whether the target or filler word was a real word (the same trial was repeated for the six pairs with each text). After the lexical decision task, they answered three yes–no comprehension questions. After finishing the comprehension questions of the last text, they answered the English-reading proficiency test within 30 minutes. The whole procedure lasted for about 60 minutes.

3.3.2.4 Scoring and Analysis

The correct response times deviating more than 2.5 SD from the mean of an individual participant were replaced with the value of the mean plus 2.5 SD. This resulted in the replacement of 1.23% of the data (7 out of 571 data samples). The correct response times and rates were averaged in each condition across the proficiency groups and three critical pairs. Addressing the RQ, 2 (proficiency: higher, lower) × 3 (pair: major-sub, detail-sub, control-sub) two-way mixed ANOVAs were performed for the correct response times and rates, with proficiency as a between-participant factor and pair as a within-participant factor.

3.3.3 Results

3.3.3.1 English-Reading Proficiency Test

Table 3.27 illustrates the descriptive statistics of the English-reading proficiency test (Cronbach's α = .80). The participants were divided into a higher-proficiency group (n = 16) and lower-proficiency group (n = 18) based on the median of their scores. A t test showed that the scores significantly differed by proficiency group, $t(32) = 7.26$, $p < .001$, $d = 2.49$. The *t*-test result confirmed that English-reading proficiency was higher in the higher-proficiency group than in the lower-proficiency group.

The higher-proficiency group was estimated to have an English-reading proficiency between the second and pre-first grades of the EIKEN test (levels B1 and B2 on the CEFR; Dunlea, n.d). Their correct answer rates exceeded 60% for the second grade test but failed to exceed 70% for the pre-first grade test. On the other hand, the English-reading proficiency of the lower-proficiency group fell between the pre-second and second grades (levels A2 and B1 on the CEFR; Dunlea, n.d). Their correct answer rates were lower than 70% for the pre-first grade test and approximately 60% for the second grade test.

Table 3.27

Descriptive Statistics of the English-Reading Proficiency Test in Experiment 2B

		Second (k = 20)		Pre-first (k = 8)		Total (k = 28)	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	16	16.25	1.48	2.88	2.09	19.13	2.53
Lower	18	11.39	3.01	0.50	1.04	11.89	3.20

3.3.3.2 Yes–No Comprehension Questions

Yes–no comprehension questions were conducted to motivate the participants to read the texts for comprehension and to confirm that they were not too difficult for them. Table 3.28 illustrates that the participants answered more than 85% of the questions correctly, regardless of their proficiency group. The high correct answer rates confirmed that all participants understood the texts sufficiently.

Table 3.28

Descriptive Statistics of the Yes–No Comprehension Questions in Experiment 2B

	n	*M*	*SD*	Max	Min	95% CI
Higher	16	.95	.05	1.00	.81	[.92, .98]
Lower	18	.87	.10	1.00	.65	[.82, .92]

3.3.3.3 Correct Response Times of the Lexical Decision Task

To address the RQ, it was examined whether the correct response times to the target words differed among the three critical pairs in the lexical decision task. Table 3.29 and Figure 3.9 illustrate the descriptive statistics of the correct response times. A 2 (proficiency: higher, lower) × 3 (pair: major-sub, detail-sub, control-sub) two-way mixed ANOVA was performed, with proficiency as a between-participant factor, and pair as a within-participant factor.

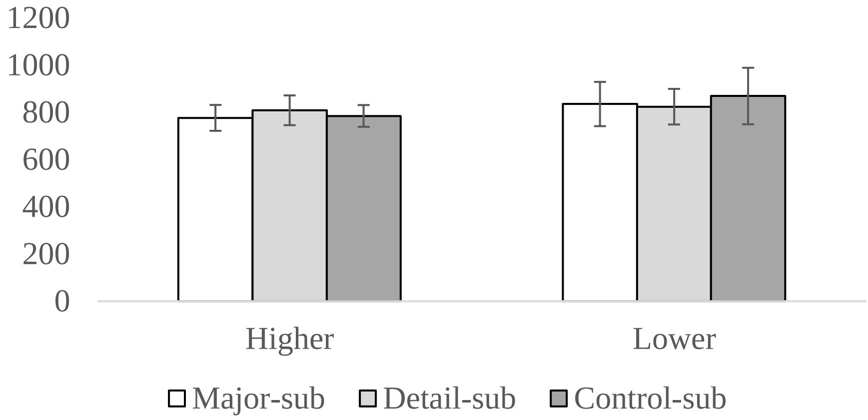

Figure 3.9. Mean correct response times (ms) of the lexical decision task (± *SEM* bars).

Table 3.29

Descriptive Statistics of the Correct Response Times (ms) *of the Lexical Decision Task*

		Major-sub		Detail-sub		Control-sub	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	16	777	219	809	252	785	185
Lower	18	836	398	824	320	869	508

The results indicated that neither the interactions nor main effects were significant (all *p*s > .10; see Table 3.30), suggesting that the participants did not make correct lexical decisions for the major-sub pairs or the detail-sub pairs more quickly, compared to correct responses to the control-sub pairs. That is, the links among the pieces of text information were not strongly activated in the participants' minds, regardless of their proficiency.

Table 3.30

Summary of the Two-Way ANOVA for the Correct Response Times of the Lexical Decision Task

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	70774.38	1.00	70774.38	0.22	.642	.01
Error (P)	10284458.39	32.00	321389.32			
			Within-participants			
Pair (P')	7481.21	1.61	4647.48	0.33	.671	.01
P × P'	20625.80	1.61	12813.18	0.92	.387	.03
Error (P')	718341.61	51.51	13945.28			

3.3.3.4 Correct Response Rates of the Lexical Decision Task

To address the RQ, correct response rates were compared among the three critical pairs. Table 3.31 and Figure 3.10 show the descriptive statistics of the correct response rates. A 2 (proficiency: higher, lower) × 3 (pair: major-sub, detail-sub, control-sub) two-way mixed ANOVA was performed for the correct response rates. The result revealed that the interaction was not significant, $F(2, 64) = 0.70$, $p = .499$, $\eta_p^2 = .02$, nor was the main effect of proficiency significant, $F(1, 32) = 0.04$, $p = .834$, $\eta_p^2 < .01$.

Table 3.31

Descriptive Statistics of the Correct Response Rates of the Lexical Decision Task

		Major-sub		Detail-sub		Control-sub	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	16	.96	.08	1.00	.00	.96	.10
Lower	18	.98	.05	.99	.05	.94	.11

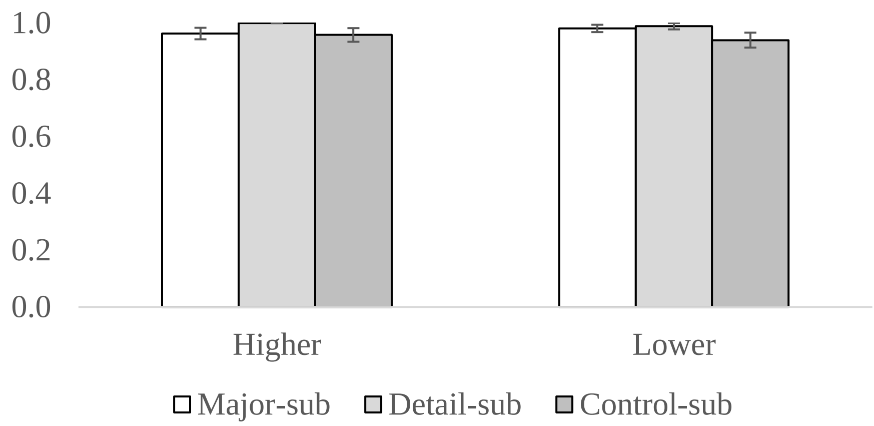

Figure 3.10. Mean correct response rates of the lexical decision task (± *SEM* bars).

On the other hand, the main effect of the pairs was significant, $F(2, 64) = 3.95$, $p = .024$, $\eta_p^2 = .11$. Multiple comparisons demonstrated that the correct response rates were significantly higher for the detail pairs than control pairs, $p = .047$, $M_{\text{diff}} = .05$, 95% CI [.00, .09]. However, there was no significant difference found between the major pairs and control pairs, $p = .545$, $M_{\text{diff}} = .02$, 95% CI [–.02, .07], or between the major pairs and detail pairs, $p = .318$, $M_{\text{diff}} = –.02$, 95% CI [–.06, .01]. Table 3.32 summarizes the results of the ANOVA.

Table 3.32

Summary of the Two-Way ANOVA for the Correct Response Rates of the Lexical Decision Task

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	< 0.01	1	< 0.01	0.04	.834	< .01
Error (P)	0.26	32	0.01			
			Within-participants			
Pair (P’)	0.04	2	0.02	3.95	.024	.11
P × P’	0.01	2	< 0.01	0.70	.499	.02
Error (P’)	0.28	64	< 0.01			

3.3.4 Discussion

The results of the correct response times and rates were contradictory in Experiment 2B, as observed in Experiment 2A. Specifically, the correct response times did not differ among the three critical pairs, indicating that the links among text information were not activated or readily available in the participants’ minds. On the other hand, while no other difference was observed among the three priming pairs, the correct response rates were higher for the detail-sub pairs than for the control-sub pairs, suggesting that the participants understood the links between the subtopics and supporting details. However, this result must be interpreted with caution because M + 1SD exceeded the maximum possible score (i.e., 1.00) regardless of critical pair or proficiency group, demonstrating a ceiling effect.

As for the readers' understanding of the links between the subtopics and supporting details, similar results were obtained in Experiments 2A and 2B. As discussed in Experiment 2A, the participants in this study might have also linked the subtopics with the supporting details while reading. Text comprehension models suggest that it is possible that they understood the incoming supporting details based on the subtopics at the beginning of the paragraphs, then summarized and integrated the supporting details into the subtopics upon finishing each paragraph (e.g., Gernsbacnher, 1990). Previous L2 research also supports this possibility, showing that Japanese university students activated subtopics during reading that were constructed from supporting details (Mori, 2015) and that readers understood themes of paragraph-long narratives (Kimura, 2013). However, they might have deactivated their understanding of the links between the subtopics and supporting details when encountering new paragraphs to ensure the necessary cognitive resources, as hypothesized in the abovementioned text comprehension models (e.g., Gernsbacnher, 1990). This deactivation was also likely to occur in this study because cognitive resources are more limited during EFL reading than L1 reading (e.g., Morishima, 2013).

On the other hand, the results concerning reader understanding of the links between the major topics and subtopics differed between the experiments. That is, while such understandings were observed in the correct response rates in Experiment 2A, they did not appear in Experiment 2B. Addressing RQ2-2, no difference in the correct response times and rates between the major-sub and control-sub pairs revealed that the participants failed to understand topic structure while reading. The different results of the correct response rates between Experiments 2A and 2B are probably due to the different priming tasks. In other words, the existence of the recognition task might have induced the participants in Experiment 2A to understand topic structure in the following two ways, although that understanding was not noticeable enough to affect the correct response times.

First, the participants' expectations of the post-reading tasks might have affected during-reading processes. The framework of the standard of coherence (e.g., van den Broek et al., 2015) assumes that readers set their standards of coherence (i.e., types and strength of coherence) in accordance with the given instructions and adapt their reading processes to achieve the expected standards. Previous research has reported that L2 readers (e.g., Nahatame, 2014; Ushiro et al., 2017) as well as L1 readers (e.g., Linderholm & van den Broek, 2002; van den Broek et al., 2001) adapt their reading processes and then generate more inferences in accordance with the given instructions.

In the case of Experiment 2A, the participants were told in advance that they would answer the recognition task as well as yes–no comprehension questions after reading. Additionally, they read the practice text and answered these tasks before the experimental session. The recognition task required the participants to refer to their text memory and to determine whether it included the target words. Hence, the participants might have set their standards of coherence at a more global level to construct more complete representations of the texts for the recognition task. This mindset might have contributed to the readers' topic structure processing while reading.

However, it should be noted that such global-level comprehension was likely to be limited during reading because of limited cognitive resources (e.g., Morishima, 2013). This might have been why the major stimuli improved correct response rates to the target words but did not result in quicker responses, compared to the control stimuli. Among the reading processes relevant to topic structure processing, the participants were likely to have difficulty with linking the major topics with the subtopics, rather than understanding these topics individually. This was supported because the correct answer rates for the yes–no comprehension questions were sufficiently high (i.e., over 85% regardless of the proficiency groups), confirming that the participants did not have difficulty with the literal comprehension of each sentence.

On the other hand, in the case of Experiment 2B, the participants were informed of the lexical decision task and practiced it, as well as the yes–no comprehension questions. The lexical decision task did not require cognitive processes related to text comprehension (i.e., it simply asked whether the target words were real words) and the yes–no comprehension questions only required literal comprehension of individual sentences. Hence, the participants might have not set their standards of coherence to a global level, and instead simply focused on the literal meaning of each sentence. Hence, the correct response rates might have not differed between the major and control pairs in Experiment 2B, although this difference was observed in Experiment 2A.

The above explanation also seems consistent with the minimalist hypothesis (McKoon & Ratcliff, 1992). The minimal hypothesis assumes that readers establish local coherence but do not build global coherence without specific or goal-directed strategic processes. In fact, past research on L2 reading processes has shown that L2 readers have difficulty linking distant sentences during reading for comprehension without specific goals or instructions (e.g., Morishima, 2013; Ushiro et al., 2017; Ushiro, Nahatame, et al., 2016, Experiment 2). In the case of the present thesis, the participants might have strategically understood topic structure during reading when instructed to answer the recognition task. On the other hand, they did not do so without such instruction, which prevented their text comprehension from progressing beyond the local level. This could explain why the participants in Experiment 2B were able to link the subtopics with the supporting details within paragraphs but failed to link the subtopics with the major topics beyond paragraphs.

Second, the priming tasks themselves affected the understandings of topic structure. In Experiment 2B, the lexical decision task did not require cognitive processes relevant to text comprehension; rather, they just asked whether the target words were real words (Jiang, 2012; McKoon & Ratcliff, 1984). On the other hand, the recognition task required the participants to refer to their text memory. This might have induced them to link the major topics with the subtopics through the reconstruction of their text memory. As discussed in Experiment 2A, higher-level processes, including topic structure processing, are usually difficult during EFL reading. Many cognitive resources are spent on lower-level processes during reading (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016) because these processes are not sufficiently proficient in L2 readers, compared to L1 readers. However, relatively more cognitive resources were available for higher-level processes in post-reading tasks after finishing the resource-demanding lower-level processes (e.g., Hosoda, 2014; Nahatame, 2013; Ushiro, Mori, et al., 2016). Combining the available cognitive resources in post-reading tasks with reference to text memory required in the recognition task, the correct response rates suggested that the recognition task induced the participants to link the major topics with the subtopics through reconstruction of text memory in Experiment 2A.

As for the participants' L2 reading proficiency, a significant Proficiency × Pair interaction was not found. This result indicates that the participants' L2 reading proficiency did not affect their linking of text information. However, the reasons for the null effect seem to differ between the understanding of the two types of links: those within and beyond paragraphs. Regarding the links between the subtopics and supporting details, the participants appeared to have understood these links regardless of their L2 reading proficiency, which is consistent with the results of Experiments 1 and 2A. This view is supported by past studies (e.g., Ushiro et al., 2009), which have suggested that L2 reading proficiency is more likely to affect literal comprehension of individual sentences than understandings of important text information. The correct answer rates of the yes–no comprehension questions were high (i.e., over 85%) for both lower- and higher-proficiency participants, confirming that both proficiency groups in Experiment 2B seemed to understand the text sufficiently. On the other hand, regarding the links between the subtopics and major topics, the correct response times and rates demonstrated that both proficiency groups likely had difficulty comprehending the links across paragraphs. This is consistent with the findings of Johns and Mayes (1990), who reported that L2 readers produced few macropropositions in the summary task, regardless of L2 reading proficiency. As in previous studies (Ushiro et al., 2009), different factors might have caused the null proficiency effect on linking within and beyond paragraphs.

3.3.5 Conclusion of Experiment 2B

The purpose of Experiment 2B was to replicate Experiment 2A and examine whether the Japanese EFL readers linked the subtopics with the major topics and supporting details during reading, using the lexical decision task instead of the recognition task.

As for the readers' understanding of the links between the subtopics and supporting details, a similar result was obtained in Experiment 2A. That is, compared with the control stimuli, the detail stimuli tended to increase the correct response rates to the target words, although the detail stimuli did not shorten the correct response times. These results indicate that the links between the subtopics and supporting details were understood but not activated or immediately available in the participants' minds. This might have been because the participants linked the subtopics with the supporting details while reading but deactivated their understanding to preserve cognitive resources for the incoming sentences. Because the results of the detail-sub pairs were similar in Experiments 2A and 2B with the different priming tasks, linking processes between the subtopics and supporting details seemed to be a general phenomenon, rather than the outcome of a specific task.

On the other hand, regarding reader understanding of the links between the major topics and subtopics, the results differed between Experiment 2A and 2B. Whereas the major stimuli did not improve the correct response times in Experiment 2A or 2B, they improved the correct response rates in Experiment 2A only. Hence, improved reader understanding of topic structure seemed to be an outcome of the recognition task. Unlike the recognition task, the lexical decision task did not require the use of cognitive processes related to text comprehension. Hence, the participants did not attempt to establish globally coherent comprehension while reading; rather, they simply focused on the literal understanding of each sentence.

In addition, the recognition task itself might have affected the participants' understanding of topic structure in Experiment 2A. Specifically, reference to text memory in the recognition task might have induced the reconstruction of their text memory, which led to an improved understanding of topic structure. On the other hand, in Experiment 2B, the participants did not have to use their text comprehension to answer the lexical decision task; thus, no improvement in the understanding of topic structure occurred.

3.4 Conclusion of Study 1

Study 1 examined whether and how Japanese EFL readers linked the major topics with subtopics in their text memory (Experiment 1) and during reading (Experiments 2A and 2B). Furthermore, such understandings were also compared to understandings of links between the subtopics and supporting details within paragraphs.

Experiment 1 investigated the memory of links among the major topics, subtopics, and supporting details in an expository text. The subtopic recall rates demonstrated that the Japanese EFL readers represented links among the above text information in their text memory and the links were retained even in their long-term memory. The subtopics were linked more robustly with the major topics than with the supporting details. This is possibly because the readers considered the major topics to be more important and related to more text information than the supporting details. The readers might have linked the text information during reading, or they might have built the links through reconstructing or reorganizing their text memory in the recall task after reading.

Experiment 2A examined whether and how Japanese EFL readers linked the subtopics with the major topics and supporting details during reading, adopting the priming paradigm. The correct response times and rates indicated that they linked the subtopics with the supporting details during reading but deactivated their understanding upon encountering new paragraphs. This is because cognitive resources available during EFL reading are limited because of the need to dedicate more to the lower-level processes required for reading (e.g., word recognition and syntactic parsing). On the other hand, as for understandings of topic structure, there are two possible explanations. First, the readers linked the major topics with the subtopics while reading, albeit to a limited extent because of restricted cognitive resources during EFL reading, as mentioned above. Second, because the recognition task required the readers to refer to their text memory, it might have induced them to reconstruct their text memory, which may have contributed to an increase in their understanding of topic structure.

Finally, Experiment 2B followed up Experiment 2A, using the lexical decision task. As for the understandings of links between the subtopics and supporting details, similar results were observed, suggesting that the readers established such links during reading and deactivated them to ensure cognitive resources for the incoming paragraphs. On the other hand, they failed to understand topic structure, unlike in Experiment 2A. This refined the first possibility in Experiment 2A: The readers might have attempted to construct more complete representations of the texts to prepare for the recognition task, which may have helped them understand topic structure. Additionally, the second possibility in Experiment 2A was confirmed because reader understanding of topic structure was observed in the recognition task but not in the lexical decision task in Experiment 2B.

The main findings of Study 1 can be summarized into three points. First, topic structure processing during EFL reading is likely to be difficult. The results of Experiment 2A suggest that although the links between the major topics and subtopics were understood, this understanding was not sufficient to be strongly activated in the readers' minds. Second, Japanese EFL readers were likely to understand topic structure through the reconstruction of text memory. This was supported by the difficulty linking the major topics and subtopics during reading found in Experiment 2A, and the memory of these links represented in text memory in Experiment 1. Third, reader understanding of topic structure was affected by the expectation of a post-reading task or its completion. This is suggested by the different results of the correct response rates in Experiments 2A and 2B.

Therefore, Study 2 will investigate how to support EFL readers' understanding of topic structure, especially during reading. By comparing the results of Experiments 2A and 2B, the effects of reading instruction (Experiment 3), and task engagement (Experiment 4) will be explored. Although Experiment 2A adopted the recognition task, adopting a task that orients readers into globally coherent comprehension seems more effective in improving reader understanding of topic structure. Therefore, the next chapter will investigate whether text-outlining instructions and participant task engagement can aid reader understanding of topic structure.

Chapter 4
Study 2: Effects of Educational Interventions on Reader Understanding of Topic Structure

4.1 Experiment 3: The Effects of Outline Instructions on Reader Understanding of Topic Structure

4.1.1 Purpose, Hypotheses, and Research Questions

The purpose of Experiment 3 is to examine whether reading instruction can prompt understanding of topic structure during reading and in post-reading tasks. The results in Experiments 2A and 2B, in particular, the correct response times, demonstrated that Japanese EFL readers had difficulty with topic structure processing while reading. These results were consistent with the findings of previous studies, which showed that the Japanese EFL readers had difficulty linking distant sentences during reading (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016). One possible source of this, as explained in the previous chapters, is the challenge of limited cognitive resources competing with lower-level processes during EFL reading. Hence, EFL readers need some support in allocating cognitive resources to topic structure processing.

In addition, the results of the correct response rates differed between Experiments 2A and 2B, suggesting that the use of the recognition task might have supported reader understanding of topic structure. One possibility is that the participants payed more attention to the overall contents of the texts during reading to answer the recognition task, in contrast to the lexical decision task. On the recognition task, the participants needed to understand as much of the text as possible so that they could refer to this text memory during the recognition task. Therefore, this experiment examined the effect of reading instruction on the participants' topic structure processing during reading as well as its memory. This is because topic structure processing during reading has been found to affect readers' text memory (e.g., Lorch & Lorch, 1985).

The framework of standards of coherence (e.g., van den Broek et al., 2015) hypothesizes that readers set their standards of coherence (i.e., the types and strength of coherence) in accordance with the given instruction, which alters reading processes and enhances text memory. In fact, several L2 studies have shown that L2 readers adapt their reading processes to given instructions (e.g., Nahatame, 2014; Ushiro et al., 2017). To support Japanese EFL readers' understanding of topic structure, outline instructions were given to the participants. The outline instructions required the readers to itemize the major topics and subtopics of the texts using different kinds of bullets (e.g., circles for major topics and dots for subtopics). Hence, this type of instruction prompts the readers to identify important information in texts (i.e., the major topics and subtopics) and organize it in accordance with topic structure. In fact, previous studies suggested that outline instructions enhance L1 readers' memory of topic structure (Lorch et al., 2013; Lorch et al., 1987). Thus, outline instructions may be able to help the Japanese EFL readers improve their understanding of topic structure. The hypotheses (Hs) and RQs addressed in this experiment were as follows:

H3-1: Japanese EFL readers have difficulty understanding topic structure during reading that is measured by self-paced reading.

H3-2: Japanese EFL readers can represent topic structure in their text memory.

RQ 3-1: Do the outline instructions support topic structure processing during EFL reading?

RQ 3-2: Do the outline instructions support memory of topic structure after EFL reading?

Topic structure processing during reading and its memory were measured based on the methods of previous studies (Hyönä & Lorch, 2004; Lorch & Lorch, 1985; Lorch et al., 2001; Murray & McGlone, 1997). It was examined whether reading times and recall rates of the subtopics differed according to the explicitness of the major topic in the text. If the participants attempt to understand topic structure while reading, this should enable them to understand the subtopics better when the major topics are explicit in earlier parts of the text, compared to when they are not. On the other hand, if the participants are able to represent topic structure in their text memory, they should recall more subtopics when the major topics are explicit in the texts than when they are not explicit.

4.1.2 Method

4.1.2.1 Participants

Twenty-one Japanese undergraduate and graduate students participated in this experiment. Their majors were varied, as follows: health and physical education, humanities and culture, human sciences, life and environmental sciences, medicine and medical sciences, science and engineering, and social and international studies. All the participants were native speakers of Japanese who had received at least six years of EFL classes as part of their formal education in Japan. None of them participated in the other experiments of this study.

Based on CEFR alignment studies (Dunlea, n.d.; ETS, 2015, 2017), their general English proficiency was estimated to range from the beginner to the upper-intermediate level (i.e., about A1 to B2 levels on the CEFR) according to self-reported scores on the TOEIC listening and reading test (range: 620–845; 601–700: $n = 1$; 801–900: $n = 1$), the TOEFL ITP test (range: 430–590; 401–500: $n = 3$; 501–600: $n = 6$), and the EIKEN test (range: 3rd–pre-1st; 3rd: $n = 7$; pre-2nd: $n = 1$; 2nd: $n = 3$; pre-1st: $n = 2$). Note that some participants did not report any of their standardized test scores or grades, whereas other participants reported all of them.

4.1.2.2 Materials

The English-reading proficiency test was the same as that used in Study 1. However, some changes were made to the texts and the yes–no comprehension questions from Study 1.

Experimental and practice texts

A total of eight texts that represented topic structure were collected and revised from previous studies (Coté et al., 1998; Goldman et al., 1995; Hidi & Baird, 1988; Kintsch, 1990; Kobayashi, 2002; Lorch, 1993; McNamara, Kintsch, Songer, & Kintsch, 1996; Ohlhausen & Roller, 1988). Each experimental text included one introductory sentence, one major topic that summarized the whole text, one subtopic for each of the three body paragraphs, and supporting details (see Appendix 3). Each text began with an introductory sentence, followed by a sentence introducing the major topic, to prevent the major topic from appearing at the beginning of the texts. In the body of the text, the subtopics were followed by the supporting details. To examine reader understanding of topic structure, two conditions were set for each of the experimental texts: The experimental texts were presented to the participants either with or without the major topics (i.e., *explicit* and *non-explicit conditions*, respectively). It was examined whether the participants understood topic structure during reading and represented it in their text memory by comparing reading processes and memory of subtopics in the explicit and non-explicit conditions.

Because the focus of this experiment was discourse-level comprehension, readability (according to the FKGL) and word length were controlled to avoid the possible effects of differences between the texts. Table 4.1 displays these features of the texts. In addition, low-frequency words at Level 5 or above of the *JACET 8000 list* (JACET, 2003) were paraphrased using high-frequency words at Level 4 or below. After the above revision, a native speaker of English proofread the texts. Moreover, in addition to the experimental texts, two practice texts with similar topic structure, text length, and readability were also created. The practice texts were presented in the explicit condition alone.

Table 4.1

Length and Readability of the Experimental Texts in Experiment 3

Text	Words	FKGL	Content
Chimpanzee	166	7.5	Chimpanzees' interesting lifestyles
Energy	171	8.1	Major problems of energy production
Film	165	7.8	Film genres that appeared in the 1930s
Heart disease	173	7.9	Ways to deal with heart diseases
Inventor	165	8.0	Common characteristics among great inventors
Nationalism	170	8.7	Factors that affect nationalism
Peru	179	7.0	Characteristics of Peru as a country
Support	174	8.1	Problems of supporting developing countries

Note. The FKGL was provided by Microsoft Word 2010's readability measurement tools.

Pilot study

A pilot study was conducted to confirm whether the experimental texts had identifiable topic structures. A total of 11 Japanese university students participated in the pilot study. They read the experimental texts in the explicit condition and rated how important each sentence was in understanding the entire text, using a 5-point Likert scale (i.e., 1 = *not important at all*, 2 = *not important*, 3 = *neither unimportant nor important*, 4 = *important*, 5 = *very important*). None of them participated in the main study. Table 4.2 illustrates the descriptive statistics of the importance ratings of each text.

Table 4.2

Descriptive Statistics of the Importance Ratings in the Pilot Study of Experiment 3

	Major topic		Subtopic		Supporting detail	
Text	*M*	*SD*	*M*	*SD*	*M*	*SD*
Chimpanzee	4.73	0.65	4.15	0.56	2.65	0.73
Energy	4.82	0.40	3.97	0.46	2.41	0.52
Film	4.64	0.50	3.88	0.40	2.64	0.80
Heart disease	4.55	0.69	3.88	0.54	2.37	0.75
Inventor	4.82	0.60	4.09	0.62	2.27	0.64
Nationalism	4.64	0.50	3.94	0.42	2.36	0.70
Peru	4.64	0.67	4.00	0.39	2.39	0.61
Support	4.82	0.40	4.06	0.47	2.58	0.56

To confirm the validity of topic structure of the texts, it was examined whether the importance ratings were higher for the major topics and lower for the supporting details than for the subtopics. For this end, a one-way repeated ANOVA was conducted, with the information as an independent variable and the ratings as a dependent variable. This analysis yielded a significant effect for information on the ratings, $F(1.23, 12.28) = 71.58$, $p < .001$, $\eta_p^2 = .88$. Multiple comparisons found a significant difference between the major topics and subtopics, $M_{\text{diff}} = 0.71$, $p = .002$, 95% CI [0.29, 1.13], and also between the subtopics and supporting details, $M_{\text{diff}} = 1.54$, $p < .001$, 95% CI [1.10, 1.98]. These results confirmed that the experimental texts included clear topic structures. Table 4.3 summarizes the statistical results.

Table 4.3

Summary of the One-Way ANOVA for the Importance Ratings in the Piot Study of Experiment 3

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Information (I)	29.06	1.23	23.66	71.58	< .001	.88
Error (Individual difference)	2.31	10.00	0.23			
Error (I)	4.06	12.28	0.33			

Material sets

Four material sets were created following the Latin square procedure. In each set, two of the eight experimental texts were assigned to each of the four conditions crossing major-topic explicitness (i.e., explicit or non-explicit) and reading condition (i.e., standard or instruction). Across the material sets, each text was presented once in one of the four conditions.

Yes–no comprehension questions

As in Experiments 2A and 2B, one yes–no comprehension question was created for each text to motivate the participants to read for comprehension and to confirm whether literal comprehension of the texts was possible for the participants. Each question queried literal understanding of a sentence representing a supporting detail. The correct answer was *yes* for half of the texts (k = 4) and *no* for the other half of the texts (k = 4).

4.1.2.3 Procedure

The experiment was conducted with each participant individually. Before it began, the experimenter explained the general purpose and procedure of the research and obtained the informed consent of each participant. The participants then filled out the profile sheet including the self-report of their standardized test scores. The whole procedure of the main session lasted for approximately 90 minutes. The main session consisted of the reading session, cued recall task, and English-reading proficiency test. Figure 4.1 illustrates the procedure of the main session.

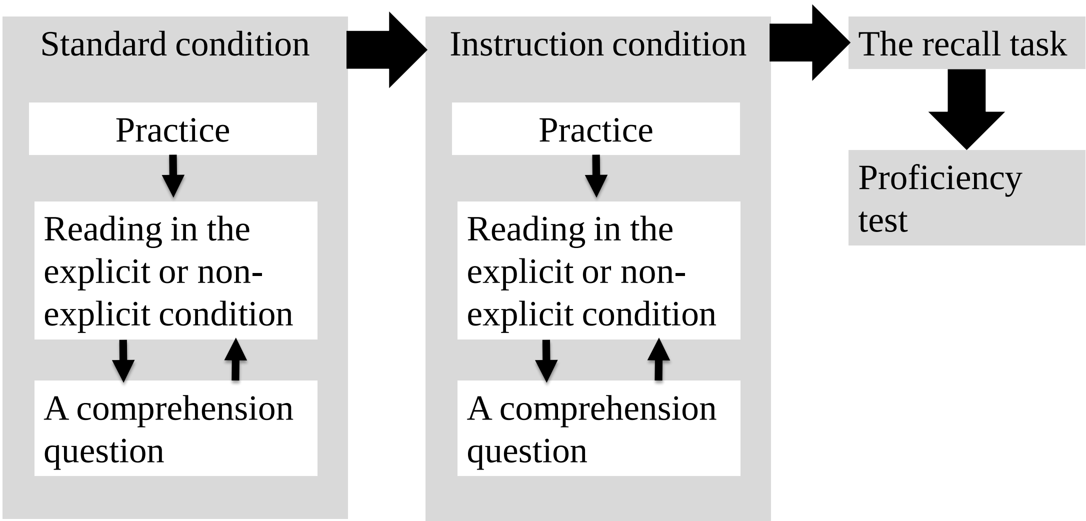

Figure 4.1. Procedure of Experiment 3.

In the reading session, eight experimental texts were presented on a computer screen in random order using SuperLab 5.0. software (Cedrus, US). Half of the texts were presented in the explicit condition and the other half in the non-explicit condition. The participants read the texts sentence-by-sentence at their own pace. They pressed the "yes" key on the response pad (RB-740 model, Cedrus, US) to replace the focal sentence with the next one. The intervals between key presses were recorded as the reading times. When the participants finished reading each text, a warning ("???") appeared on the screen for 700 ms to inform the participants of the yes–no comprehension question. When pressing the "yes" or "no" key to answer the question, the participants received accuracy feedback for 2,000 ms. This trial was repeated for each of the eight texts.

To examine the instruction effect within participants, all participants read the first four texts without instruction for a specific task (i.e., *standard condition*) and the latter four texts after receiving the outline instructions (i.e., *instruction condition*). Following the methods used in previous studies that investigated instruction effects on reading comprehension (e.g., Kimura, 2014, 2015a; Nahatame, 2014; Ushiro et al., 2017), the order of the comprehension and instruction conditions was fixed. This is because the participants might not read the texts for comprehension after reading for the outline task.

In the standard condition, the participants were notified of the yes–no comprehension questions in advance and read the texts to answer them. On the other hand, in the instruction condition, the participants were instructed to read the texts to write outlines afterward, as well as to answer the comprehension questions. Based on the reading instructions given in Lorch et al. (2013), the following instructions were given to the readers in the present experiment:

> The purpose of reading this time is to write an outline of each text. For this end, please read the texts so that you can write outlines of each text after reading the other four texts. Please prepare by itemizing the major topics that are the most important to understand the whole text as well as the next important subtopics, using circles ("●") for the major topics and dots ("・") for subtopics.

The participants read a practice text at the beginning of each reading condition to familiarize themselves with reading in these conditions. They read the practice texts and answered the yes–no comprehension questions. In addition, at the end of the practice session for the instruction condition, the participants were presented with a sample outline of the practice text. Table 4.4 displays the sample outline.

Table 4.4

Sample Outline Presented in the Practice Session for the Instruction Condition

● There are an increasing number of acts for preserving the environment.
・ These acts include those for air quality conservation.
・ These acts include those for water conservation.
・ These acts include those for prohibiting the throwing away of garbage.

Note. The sample outline was presented in Japanese in the experiment and translated into English by the author.

After the reading session, the participants completed the written recall task for the eight experimental texts. The introductory sentence of each text was given as a recall cue, and they were allowed to answer from any of the texts they remembered. The time was not limited so that the participants had sufficient time to write down all they could remember. After that, the participants answered the English-reading proficiency test within 30 minutes.

4.1.2.4 Scoring

Before calculating the reading times, the outliers deviating more than 2.5 *SD*s from the mean of an individual participant in each condition were replaced with the value of the mean plus 2.5 *SD*s. This resulted in the replacement of 5.49% of the data (293 out of 5336 data samples). To control for the effect of sentence length, the reading time for each sentence was divided by the number of syllables. The average reading time for each sentence was calculated for both the standard and instruction conditions. Additionally, the average reading time for the subtopics was calculated in each of the explicitness conditions and reading conditions.

Before the scoring of the recall protocols, two graduate students who majored in

English education divided three texts (more than 30% of eight texts) into IUs based on Ikeno's (1996) criteria. The inter-rater agreement rate as the scoring reliability was 96.53%. After the raters discussed and resolved all disagreements, one of them divided the other five texts into IUs. Based on the IU division, the author scored the recall protocols from 15% of the participants (n = 4) with a graduate school student majoring in English education. In addition, the author scored the recall protocols from another 15% of the participants with another graduate school student of the same major. The inter-rater agreement rates as the scoring reliability were 93.67% and 89.45% for each pair. All discrepancies were resolved through discussion. Based on the criteria, the author scored the rest of the recall protocols.

Next, the total recall rates were calculated in the standard and instruction conditions. Moreover, the subtopic recall rates were calculated in each condition crossing the explicitness conditions of the major topics and the reading conditions. Because the introductory sentences were presented as the recall cues, and the participants did not read the major topics in the non-explicit condition, these sentences were excluded from the total recall rates. Because the number of IUs differed among the texts assigned to each condition, the recall rates were normalized by arcsine transformation before the statistical tests were conducted.

4.1.2.5 Analysis

To examine general effects of the outline instructions on reading processes and text memory, 2 (proficiency: higher, lower) × 2 (reading: standard, instruction) two-way mixed ANOVAs were performed for the average reading times and the total recall rates. In the analyses, reading proficiency was considered a between-participant factor and reading conditions a within-participant factor. Further, to analyze the instruction effect on topic structure processing and memory, 2 (proficiency: higher, lower) × 2 (major topic: explicit, non-explicit) × 2 (reading: standard, instruction) three-way mixed ANOVAs were carried out for the subtopic reading times and recall rates. In addition to the factors of reading proficiency and condition mentioned above, major-topic explicitness was included as a within-participant factor. These analyses were conducted to address RQs 3-1 and 3-2: whether the participants understood the subtopics faster and represented them in their text memory more robustly in the explicit condition than in the non-explicit condition.

Further, the correlation between topic structure processing during reading and memory of topic structure were analyzed. For this end, the differences in subtopic reading times and recall rates were calculated for the explicit and non-explicit conditions. Topic structure processing and memory of topic structure were assumed to appear in the shorter subtopic reading times and the higher subtopic recall rates in the explicit than in the non-explicit condition. Hence, the differences between the explicitness conditions were assumed to reflect the degree to which the participants understood topic structure during reading and represented it in their text memory. After calculating these values, Spearman's rank correlation coefficients were calculated for the differences in the standard and instruction conditions.

4.1.3 Results

4.1.3.1 English-Reading Proficiency Test

Table 4.5 shows the descriptive statistics for the English-reading proficiency test (Cronbach's α = .79). The participants were divided into a higher-proficiency group (n = 11) and lower-proficiency group (n = 10) based on the median. A t test confirmed that the test scores were significantly different between the groups, $t(19) = 6.47$, $p < .001$, $d = 2.83$.

The English-reading proficiency of the higher-proficiency group was approximately between the second and pre-first grades of the EIKEN test (levels B1 and B2 on the CEFR; Dunlea, n.d.). Their correct answer rates surpassed 60% for the second grade test but they did not exceed 70% for the pre-first grade test. On the other hand, the English-reading proficiency of the lower-proficiency group was between the pre-second and second grades (levels A2 and B1 on the CEFR; Dunlea, n.d.). Their correct answer rates were lower than 70% for the pre-first grade test but close to 60% for the second grade test.

Table 4.5

Descriptive Statistics of the English-Reading Proficiency Test in Experiment 3

		Second (k = 20)		Pre-first (k = 8)		Total (k = 28)	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	11	16.90	2.39	2.55	1.75	19.45	2.58
Lower	10	10.90	2.64	1.00	1.33	11.90	2.77

4.1.3.2 Yes–No Comprehension Questions

Yes–no comprehension questions were created to motivate the participants to read the texts for comprehension. Table 4.6 shows the descriptive statistics for the yes–no comprehension questions. As the table shows, the correct answer rates were higher than 85%, regardless of the participants' English-reading proficiency and the reading conditions. This confirmed that both proficiency groups understood the texts without difficulty.

Table 4.6

Descriptive Statistics of the Yes–No Comprehension Questions in Experiment 3

		Standard			Instruction		
	n	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Higher	11	.91	[.82, .99]	.13	.89	[.80, .97]	.13
Lower	10	.98	[.92, .103]	.08	.87	[.69, 1.04]	.25

4.1.3.3 Reading Times

Table 4.7 and Figure 4.2 show the descriptive statistics of the average reading times. To investigate the instruction effect on reading processes in general, a 2 (proficiency: higher, lower) × 2 (reading: standard, instruction) two-way mixed ANOVA was carried out for the average reading times. The interaction was not significant, $F(1, 19) = 0.47$, $p = .503$, $\eta_p^2 = .02$. However, a main effect of proficiency was significant, $F(1, 19) = 5.54$, $p = .030$, $\eta_p^2 = .23$, indicating that the higher-proficiency readers read faster than the lower-proficiency readers. Moreover, the main effect of reading was also significant, $F(1, 19) = 23.72$, $p < .001$, $\eta_p^2 = .56$, showing that the participants read the texts more carefully in the instruction condition than in the standard condition. Table 4.8 summarizes the results of the two-way ANOVA.

Table 4.7

Descriptive Statistics of the Average Reading Times (ms) *per Syllable*

		Standard			Instruction		
	n	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Higher	11	403	[299, 506]	128	449	[351, 546]	122
Lower	10	557	[448, 666]	198	618	[516, 721]	184

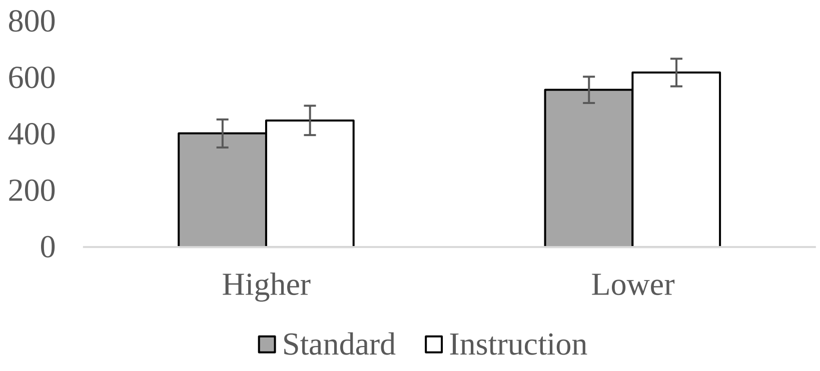

Figure 4.2. Mean of the average reading times (ms) per syllable (± *SEM* bars).

Table 4.8

Summary of the Two-Way ANOVA for the Average Reading Times

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	275450.57	1	275450.57	5.54	.030	.23
Error (P)	944923.52	19	49732.82			
			Within-participants			
Reading (R)	30369.22	1	30369.22	23.72	< .001	.56
P × R	596.60	1	596.60	0.47	.503	.02
Error (R)	24322.05	19	1280.11			

Table 4.9 and Figure 4.3 show the descriptive statistics for the subtopic reading times. To address the RQs, a 2 (proficiency: higher, lower) × 2 (major topic: explicit, non-explicit) × 2 (reading: standard, instruction) three-way mixed ANOVA was performed. A main effect for reading was significant, $F(1, 19) = 5.61$, $p = .029$, $\eta_p^2 = .23$, indicating that the participants read the subtopics more attentively when reading for the outline task than for comprehension. However, none of the other interactions or main effects were significant (all *p*s > .05).

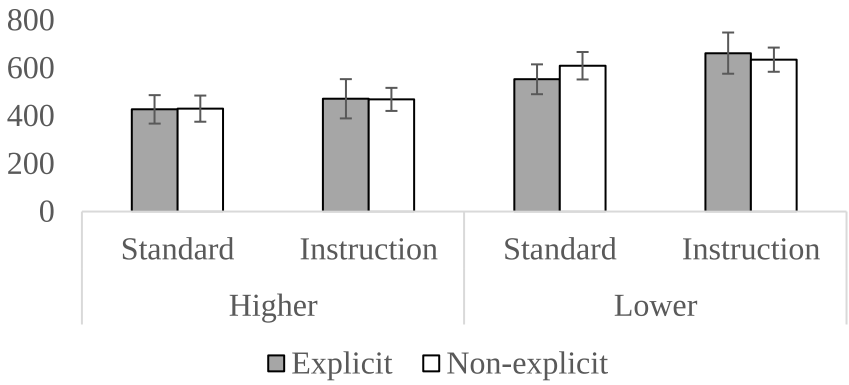

Figure 4.3. Mean of the subtopic reading times (ms) per syllable (± *SEM* bars).

Table 4.9

Descriptive Statistics of the Subtopic Reading Times (ms) *per Syllable*

		Major-topic explicit			Major-topic non-explicit		
	n	*M*	95% CI	*SD*	*M*	95% CI	*SD*
				Standard condition			
Higher	11	427	[303, 552]	162	430	[316, 545]	123
Lower	10	553	[422, 684]	230	610	[490, 730]	229
				Instruction condition			
Higher	11	471	[300, 643]	200	469	[368, 570]	127
Lower	10	662	[482, 843]	335	635	[529, 741]	190

Table 4.10 summarizes the ANOVA results. These results showed that the subtopic reading times did not differ according to the explicitness of the major topic, regardless of the proficiency group or reading condition. That is, both proficiency groups had difficulty understanding the links between the major topics and subtopics during reading, regardless of whether they were reading for comprehension or outlining.

Table 4.10

Summary of the Three-Way ANOVA for the Subtopic Reading Times

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	574121.22	1	574121.22	4.25	.053	.18
Error (P)	2563937.70	19	134944.08			
			Within-participants			
Explicitness (E)	1152.60	1	1152.60	0.09	.768	.01
E × P	1095.76	1	1095.76	0.08	.774	< .01
Error (E)	245625.11	19	12927.64			
Reading (R)	62044.07	1	62044.07	5.61	.029	.23
R × P	3496.49	1	3496.49	0.32	.581	.02
Error (R)	210202.54	19	11063.29			
E × R	10548.89	1	10548.89	0.84	.372	.04
E × R × P	8080.96	1	8080.96	0.64	.433	.03
Error (E × R)	239782.00	19	12620.11			

4.1.3.4 Recall Rates

Table 4.11 and Figure 4.4 illustrate the descriptive statistics of the total recall rates. To clarify the instruction effect on general text memory, a 2 (proficiency: higher, lower) × 2 (reading: standard, instruction) two-way mixed ANOVA was conducted. The result did not yield a significant interaction, $F(1, 19) = 0.87$, $p = .363$, $\eta_p^2 = .04$, or a significant main effect for proficiency, $F(1, 19) = 0.54$, $p = .470$, $\eta_p^2 = .03$. However, a significant main effect for reading was observed, $F(1, 19) = 6.47$, $p = .020$, $\eta_p^2 = .25$, revealing that the subtopic recall rates were higher in the instruction condition than in the standard condition. This difference suggests that participants' text memory was enhanced when the outline instructions were given. Table 4.12 summarizes the ANOVA results.

Table 4.11

Descriptive Statistics of the Total Recall Rates Normalized by Arcsine Transformation in Experiment 3

		Standard			Instruction		
	n	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Higher	11	29.84	[24.10, 35.59]	6.58	32.24	[25.66, 38.82]	7.70
Lower	10	25.51	[19.48, 31.53]	11.26	30.67	[23.77, 37.57]	12.79

Table 4.12

Summary of the Two-Way ANOVA for the Total Recall Rates in Experiment 3

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
Between-participants						
Proficiency (P)	91.34	1	91.34	0.54	.470	.03
Error (P)	3200.05	19	168.42			
Within-participants						
Reading (R)	149.63	1	149.63	6.47	.020	.25
P × R	20.09	1	20.09	0.87	.363	.04
Error (R)	439.34	19	23.12			

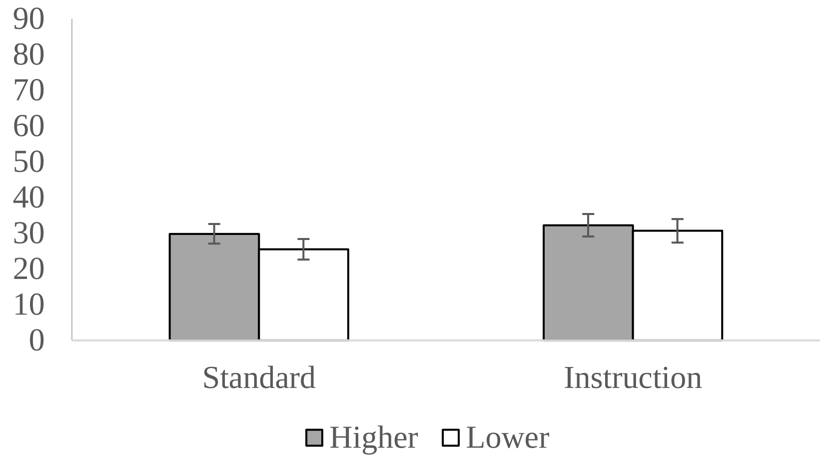

Figure 4.4. Mean of the total recall rates normalized by arcsine transformation in Experiment 3 (± *SEM* bars).

Table 4.13 and Figure 4.5 present the descriptive statistics for the subtopic recall rates. To address the RQs, a 2 (proficiency: higher, lower) × 2 (major topic: explicit, non-explicit) × 2 (reading: standard, instruction) three-way mixed ANOVA was performed for the subtopic recall rates. The result yielded a significant main effect for reading, $F(1, 19) = 19.19$, $p < .001$, $\eta_p^2 = .50$, although none of the other main effects or interactions were significant (all ps > .10). The subtopic recall rates were significantly higher in the instruction condition than in the standard condition, suggesting that the outline instructions enhanced the memory of the subtopics. However, because the memory of the subtopics did not differ according to the explicitness of the major topics, the links between the major topics and subtopics were not represented in participants' text memory. Table 4.14 summarizes the results of the three-way ANOVA.

Table 4.13

Descriptive Statistics of the Subtopic Recall Rates Normalized by Arcsine Transformation in Experiment 3

		Major-topic explicit			Major-topic non-explicit		
	n	*M*	95% CI	*SD*	*M*	95% CI	*SD*
			Standard condition				
Higher	11	30.24	[19.13, 41.35]	17.33	26.23	[18.07, 34.40]	13.94
Lower	10	23.62	[11.96, 35.27]	17.91	20.12	[11.56, 28.68]	11.71
			Instruction condition				
Higher	11	33.61	[22.68, 44.54]	20.15	38.91	[32.14, 45.67]	9.31
Lower	10	30.12	[18.65, 41.58]	13.50	38.33	[31.24, 45.43]	12.10

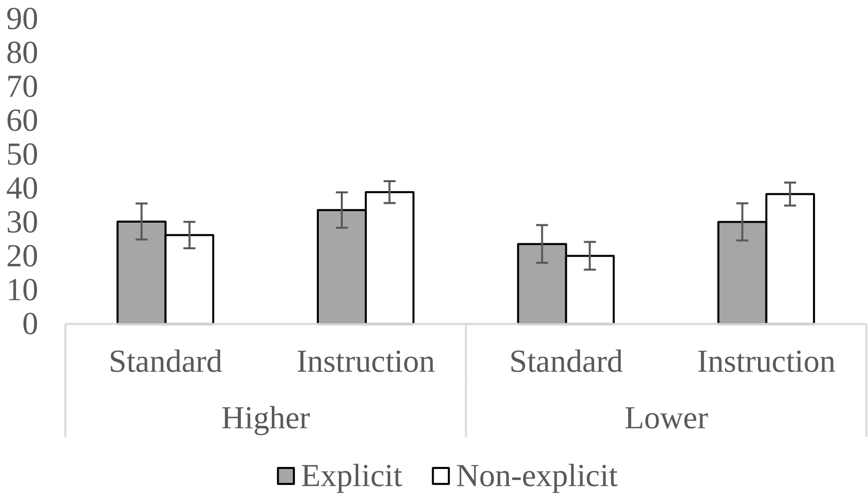

Figure 4.5. Mean of the subtopic recall rates normalized by arcsine transformation in Experiment 3 (± *SEM* bars).

Table 4.14

Summary of the Three-Way ANOVA for the Subtopic Recall Rates in Experiment 3

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	369.92	1	369.92	0.96	.340	.05
Error (P)	7326.64	19	385.61			
			Within-participants			
Explicitness (E)	47.27	1	47.27	0.33	.573	.02
E × P	15.45	1	15.45	0.11	.747	.01
Error (E)	2732.24	19	143.80			
Reading (R)	2174.47	1	2174.47	19.19	< .001	.50
R × P	98.39	1	98.39	0.87	.363	.04
Error (R)	2152.50	19	113.29			
E × R	578.39	1	578.39	2.32	.144	.11
E × R × P	7.57	1	7.57	0.03	.864	< .01
Error (E × R)	4741.63	19	249.56			

4.1.3.5 Relation of Reading Times and Recall Rates of the Subtopics

To explore the relations between topic structure processing during reading and memory of topic structure, Spearman's rank correlation coefficients were calculated for the standard and instruction conditions. The results demonstrated only weak correlations in the standard condition, $r = –.20$, $p = .382$, and instruction condition, $r = –.06$, $p = .794$. These weak correlations indicate that topic structure processing during reading and memory of topic structure are not strongly related.

4.1.4 Discussion

Topic structure processing during reading without specific reading instructions (H3-1)

The subtopic reading times did not significantly differ according to the explicitness of the major topic in the standard condition. This result indicates that the participants failed to understand topic structure during reading, which supports H3-1. The readers' difficulty with topic structure processing found in this experiment is consistent with the findings of Experiments 2A and 2B. In Experiments 2A and 2B, the correct response times indicated that links between the major topics and subtopics were not activated or readily available in the participants' minds. Moreover, the difficulty with topic structure processing was also consistent with some theoretical frameworks. In terms of the minimalist hypothesis (McKoon & Ratcliff, 1992), the participants might not have constructed globally coherent comprehension such as understanding of topic structure without specific reading instructions that would trigger strategic reading processes. Furthermore, in terms of standards of coherence (e.g., van den Broek et al., 2015), the participants might have set their standards of coherence at the sentence level because they were only required to answer yes–no comprehension questions for the supporting details in the standard condition.

As discussed in Experiments 2A and 2B, one possibility is that the participants' insufficient lower-level processing abilities (e.g., word recognition and syntactic parsing) might have depleted their cognitive resources for higher-level processes (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016, Experiment 2; Yoshida, 2003), such as topic structure processing. In addition, a methodological factor might have also limited the participants' cognitive resources. Specifically, when the participants had difficulty with topic structure processing in Experiments 2A, 2B, and 3, the texts were presented sentence by sentence, which prevented the participants from looking back at prior sentences during reading.

To link two pieces of text information, they must be activated simultaneously in the readers' minds (e.g., Kintsch, 1998). In the case of linking distant information, readers must retrieve earlier text information from their long-term memory (e.g., van der Schoot et al., 2012). For example, L1 readers looked back at headings (i.e., corresponding to the major topics) when they were attempting to understand topic structure while reading (Hyönä et al., 2002; Hyönä & Nurminen, 2006). These text references likely enabled them to reactivate their memory of the major topics to link them with the subtopics. On the other hand, the participants were not permitted to refer back to previous sentences in Experiments 2A, 2B, or 3. Thus, they had to keep the major topics activated in their minds during reading to link them with the subtopics. This might have made it more difficult to link the major topics with subtopics during reading, combined with the limited cognitive resources. Hence, it is necessary to examine topic structure processing in a situation where the participants are allowed to refer back to prior sentences. This will be further explored in the next experiment.

Memory of topic structure in the post-reading task without specific reading instructions (H3-2)

In the standard condition, the subtopic recall rates did not differ between the explicit and non-explicit conditions. This result indicates that the participants failed to represent topic structure in their text memory, which rejects H3-2. One possible reason is that the participants failed to process topic structure while reading, as discussed concerning H3-1. Past research has indicated that readers understand topic structure during reading, which helps them to retrieve text memory (e.g., Hyönä & Lorch, 2004; Lorch & Lorch, 1985). However, because the participants were unable to understand topic structure while reading, they were not able to use it in text memory retrieval.

Furthermore, other possibilities might be relevant to the poor memory of topic structure. Because topic structure processing and memory were only weakly correlated, cognitive processes in the post-reading task (e.g., reconstruction of text memory) as well as during-reading comprehension were related to memory of topic structure. Combined with poor memory of topic structure, the participants were likely to fail to understand topic structure not only during reading but also through reconstruction of text memory in the post-reading task. Compared with the participants' robust memory of topic structure that was found in Experiment 1, it is possible that the participants were not able to understand topic structure independently in Experiment 3. Specifically, the participants recalled topic structure when major cues were given in Experiment 1, whereas the participants in this experiment did not recall topic structure without such recall cues. Thus, giving the major topics as recall cues seemed to be effective scaffolding for the participants to link the major topics with the subtopics through the reconstruction of text memory. This might have been because the major topics are directly related to the subtopics in hierarchical topic structure (e.g., Britton, 1994; van Dijk & Kintsch, 1983). This possibility will be addressed in further detail in the discussion of RQ 3-2.

The effect of the outline instructions on EFL reading processes (RQ3-1)

The result revealed that the average reading times were significantly longer in the instruction condition than in the standard condition, regardless of the proficiency group. This suggests the effect of the outline instructions on EFL reading processes in general. That is, the outline instructions guided the participants to read the texts more attentively than they did without such instruction. This is consistent with the findings of previous studies, which observed more attentive reading with specific reading instructions (e.g., Ushiro et al., 2017). Hence, the participants might have adapted their reading processes in accordance with the outline instructions.

However, addressing RQ 3-1, the results demonstrate that the subtopic reading times did not differ according to the explicitness of the major topics, regardless of the reading condition, indicating that the participants failed to link the subtopics with major topic while reading, with or without the outline instructions. The theoretical framework of standards of coherence (e.g., van den Broek et al., 2015) assumes that readers set a reading goal in accordance with the given instructions and adapt their reading processes to achieve that goal. Hence, it is possible that the participants failed somewhere in these processes, leading to the ineffectiveness of the outline instructions.

It is possible that the participants did not attempt to change their reading processes when the outline instructions were given. When instructions failed to change L2 reading processes noticeably in previous studies (e.g., Kimura, 2014, 2015a; Horiba, 2000, 2013), the researchers discussed a mismatch between the readers' linguistic proficiency and the cognitive load of the instructions. That is, L2 readers consume much of their cognitive resources during lower-level cognitive processes (e.g., word recognition and syntactic parsing) because their linguistic proficiency is low compared to L1 readers. The outline instructions may have been too resource-demanding for the participants because it required them to distinguish between three hierarchical levels in the texts, that is, the major topics, subtopics, and supporting details (Lorch et al., 2013; Lorch et al., 1987). To complete the task, they participants needed to select important text information to include in their outline while rejecting other non-essential information. Once they had included all the necessary information in their outlines, they had to relate the major topics to their subtopics. However, it might have been too resource-demanding for the participants to engage in these processes at the same time while reading because their cognitive resources were being mostly dedicated to lower-level processes, leaving few resources for the higher-level processes required by this task.

However, the first possibility seemed to be rejected by the increase in the average reading times from the standard to the instruction condition. This indicates that the participants attempted to adapt their reading processes to the outline instructions. Hence, it is also possible that simply giving the outline instructions did not sufficiently guide the participants into topic structure processing. Because the participants did not actually write the outlines while reading, they might not have been fully motivated to engage in the reading process as the outline instructions intended. This might be explained by activity theory (e.g., Stull & Mayes, 2007), which assumes that readers achieve deep text comprehension, such as selection and organization of relevant information, by engaging in productive tasks. Writing outlines requires that the reader select important information from a text and organize the information into major topics and subtopics. Hence, to facilitate these cognitive processes, it might be necessary for the readers to complete the outline task themselves, rather than just give them the instructions to read for that task.

The effect of the outline instructions on text memory after EFL reading (RQ3-2)

The results show that the total recall rates were significantly higher for the instruction condition than the standard condition, suggesting that the outline instructions enhanced the participants' overall text memory. This result is consistent with the results concerning average reading times; specifically, when the outline instructions were given, the participants read more carefully. The finding of enhanced text memory is also consistent with the findings of previous studies that showed that readers' text memory was enhanced with reading instructions (e.g., van den Broek et al., 2001).

Addressing RQ 3-2, the results demonstrated that subtopic recall rates did not differ according to the explicitness of the major topics in the standard or instruction conditions. This shows that the participants had difficulty representing topic structure in their text memory in the standard condition, and that the outline instructions failed to relieve this difficulty. This is consistent with the result concerning subtopic reading times, which showed that the participants failed to link the subtopics with major topics during reading. Previous studies have shown that readers represent topic structure in their evolving representations of texts and that structure helps them with text memory retrieval (e.g., Lorch & Lorch, 1985). However, in this study, the outline instructions failed to support the participants' topic structure processing while reading and their memory of topic structure.

In addition, other reasons might be related to the ineffectiveness of the outline instructions on memory of topic structure because the memory was not significantly correlated with topic structure processing during reading in the instruction condition. A similar tendency has also been observed in previous L2 studies. Horiba (2013) reported weak correlations between the think-aloud comments (e.g., coherence-building inference) and recall rates in reading for specific reading instructions, suggesting that the relation between L2 reading processes and products is not straightforward. One possible cognitive process linking during-reading processes and text memory is the reconstruction of the memory in post-reading tasks. For example, although the Japanese university students failed to maintain coherence between distant sentences during reading (Ushiro, Nahatame, et al., 2016, Experiment 2), such readers maintained coherence through reconstructing their text memory in the recall task (Ushiro, Mori, et al., 2016). As suggested by these previous studies, during-reading processes do not necessarily lead to the construction of text memory. Therefore, combining the ineffectiveness of the outline instructions on topic structure memory with the relevance of text memory reconstruction to global coherence in text memory, the outline instructions might have failed to affect text memory reconstruction in the recall task, as well as topic structure processing during reading. Elaborating the discussion concerning H3-2, the above view suggests that the participants were not able to independently (without cues) understand topic structure in the post-reading task, even when they were given specific reading instructions.

4.1.5 Conclusion of Experiment 3

The purpose of Experiment 3 was to investigate whether the outline instructions helped the Japanese EFL readers understand topic structure during reading and represent topic structure in their text memory. The main findings of Experiment 3 can be summarized into the four following points.

First, the participants failed to understand topic structure during reading without a specific reading instruction. This supports H3-1. One possibility is that cognitive resources available for topic structure processing during reading were mostly tied into lower-level cognitive processes, as discussed in Experiments 2A and 2B. Moreover, in terms of the methodology, the sentence-by-sentence, self-paced reading might have made topic structure processing too resource-demanding. This methodology required the participants to keep the major topics activated in their minds to link them with the subtopics, not allowing the participants to refer back to the text and reactivate their memory of the major topics. To explore the effect of the methodology, the entire texts will be presented at once instead of sentence-by-sentence in the next experiment.

Second, the participants were not able to represent topic structure in their memory, either. This finding rejects H3-2 and is inconsistent with the findings of Experiment 1, in which memory of topic structure was observed. The participants failed to represent topic structure in their text memory probably because they failed at topic structure processing while reading. Moreover, they might have also failed to understand topic structure through reconstruction of text memory in the post-reading task independently. Compared to Experiment 1, where the participants represented topic structure in their text memory using the major cues, it was likely that the participants were not able to understand topic structure without the recall cues in Experiment 3.

Third, the outline instructions did not show a positive effect on topic structure processing during EFL reading (RQ3-1). Because the participants read the texts more attentively in the outline condition than in the standard condition, they attempted to adapt their reading processes in accordance with the outline instructions. However, they failed to adjust their reading processes to link the major topics with subtopics, as the outline task required. One possible reason is that just giving reading instructions was not sufficient for the EFL readers to alter their reading processes. To motivate them to establish deep text comprehension, such as understandings of topic structure, it might be necessary for them to complete the productive tasks themselves.

Finally, the outline instructions did not enhance the participants' memory of topic structure (RQ 3-2), although it did facilitate overall text memory. As with H3-2, the participants failed to understand topic structure during the post-reading task as well as during reading. In contrast to the results of Experiment 1, the participants had difficulty understanding topic structure in the post-reading tasks without recall cues, even though they were given the outline instructions.

4.2 Experiment 4: The Effects of Task Engagement on Reader Understanding of Topic Structure

4.2.1 Purpose, Hypotheses, and Research Questions

The purpose of Experiment 4 was to examine whether task engagement facilitated topic structure processing during EFL reading and the memory of it in a post-reading task. Specifically, this experiment investigated the effect of writing outlines during reading on reader understanding of topic structure. In Experiment 3, the outline instructions failed to improve reader understanding of topic structure during reading and the recall task. One possible reason is that the participants might not have been able to understand topic structure because they were simply instructed to read for the outline task, and not actually write the outline. Based on activity theory (e.g., Stull & Mayer, 2007), which assumes that deep text comprehension (e.g., selection and organization of information) is established through engaging in a productive task, it is likely necessary for L2 readers to write text outlines during reading to improve their understanding of topic structure. On the other hand, according to the cognitive load theory (e.g., Sweller, 1988), engagement in productive tasks while reading may be too resource demanding for L2 readers, especially, those who have low L2 proficiency.

In terms of methodology, several changes were made from Experiment 3. First, the entire texts were presented at the same time, rather than sentence by sentence, as in the self-paced reading method that was used in the previous experiments. This is because the self-paced reading method did not allow the participants to look back at previous sentences, which likely made it more difficult to reactivate the major topics and subtopics in their minds for linking. Moreover, as for the measurement of during-reading processes, the think-aloud method was adopted in place of the reading time method. Although the reading time method assumes that differences between conditions reflect the targeted reading processes, it was difficult to identify precisely the source of the differences, because several reading subskills contributed to the reading times (Haberlandt, 1994). In Experiment 3, although the outline instructions increased the average reading times, it was still unclear what reading processes the increase reflected. On the other hand, the think-aloud method allows researchers to assess the contents of comprehension processes at a specific point during reading (e.g., Ericson & Simon, 1993). The Hs and RQs for Experiment 4 were thus as follows:

H4-1: Japanese EFL readers have difficulty understanding topic structure during reading when entire texts are presented at once.

H4-2: Japanese EFL readers have difficulty representing topic structure in their text memory independently.

RQ 4-1: Does the outline-writing task support topic structure processing during EFL reading?

RQ 4-2: Does the outline-writing task support memory of topic structure after EFL reading?

In Experiment 4, topic structure processing during reading was evaluated using the think-aloud protocols to identify the hierarchical relations between the major topics and subtopics. Moreover, memory of topic structure was assessed by the quality of the participants' recall of text information (i.e., to what degree the outlines represented the hierarchical topic structures). These measurements were compared when the participants did not engage in any specific task during reading for comprehension (i.e., *standard condition*) and when they wrote outlines during reading (i.e., *task condition*). In Experiment 3, reader understanding of topic structure was assessed by manipulating the major-topic explicitness. However, only the texts with the explicit major topics were used in Experiment 4 to avoid putting too much cognitive load on the participants when thinking aloud and writing outlines of the texts.

4.2.2 Method

4.2.2.1 Participants

A total of 33 Japanese undergraduate and graduate students participated in this experiment (12 females and 21 males; average age = 20.97, range = 18–24). The majors of the participants were varied, as follows: health and physical education, humanities and culture, informatics, library information and media science, life and environmental sciences, medicine and medical sciences, pure and applied sciences, science and engineering, and social and international studies.

All the participants were native speakers of Japanese, and had studied EFL for more than six years in their formal education in Japan. Based on the CEFR alignment studies (Dunlea, n.d.; ETS, 2015, 2017), their general English proficiency was estimated to be from the beginner to the upper-intermediate level (i.e., levels A1 to B1 on the CEFR) according to their self-reported scores on the TOEIC listening and reading test (range: 400–760; 301–400: n = 1; 401–500: n = 4; 501–600: n = 4; 601–700: n = 3; 701–800: n = 1), the TOEFL ITP test (range: 430–530; 401–500: n = 2; 501–600: n = 2), and the EIKEN test (range: 3rd–pre-1st; 3rd: n = 4; pre-2nd: n = 5; 2nd: n = 9; pre-1st: n = 1). However, it should be noted that some participants did not report any of their TOEIC listening and reading scores, TOEFL ITP scores, or EIKEN grades, whereas other participants reported all of these test scores.

4.2.2.2 Materials

The same English-reading proficiency test as the previous experiments was used to measure L2 reading proficiency at a discourse level. As for the experimental texts, the expository texts representing topic structure were adapted from Experiment 3 with the following three changes. First, half of the eight texts (i.e., Chimpanzee, Energy problem, Inventor, and Peru) were selected from Experiment 3. Second, while each text was presented with or without the major topics (i.e., explicit and non-explicit conditions) in Experiment 3, the explicit condition alone was adopted in Experiment 4. Third, although each text used in Experiment 3 included three clusters consisting of a subtopic and its supporting details, one of the clusters was deleted from each text in Experiment 4 (see Appendix 3). These three changes were made to reduce the amount of reading, because thinking aloud and writing outlines during reading would be hard for the participants and take a long time. Table 4.15 shows the features of the experimental texts after the above revision. In addition to the experimental texts, three practice texts that had similar structures, lengths, and readability levels were prepared. These practice texts were used for the participants to practice thinking aloud and writing an outline while reading. Each of the experimental texts was paired with a yes–no comprehension question asking about a supporting detail. Half of the questions required *yes* answers ($k = 2$) and the other half required *no* answers ($k = 2$).

Table 4.15

Length and Readability of the Experimental Texts in Experiment 4

Text	Words	FKGL	Sentence
Chimpanzee	120	8.0	11
Energy problem	126	8.5	12
Inventor	124	8.7	12
Peru	124	8.0	9

Note. The FKGLs were provided by Microsoft Word 2016’s readability measurement tools.

Four counterbalancing booklets were created, following the Latin-square procedure. In each booklet, the participants read two experimental texts in the standard condition and the other two in the task condition. Moreover, each experimental text was presented in the standard condition in the two booklets and in the task condition in the other two booklets. Further, the order of the two texts in each reading condition was counterbalanced. As a result, a total of four booklets were created. It should be noted that the order of the reading conditions was fixed; that is, the task condition always followed the standard condition. If the participants first read and outlined the texts in the task condition, this way of reading would potentially affect their reading in the standard condition, which had to be natural.

4.2.2.3 Procedure

The participants were tested individually in a single session that lasted approximately 90 minutes. At the beginning of the experiment, the experimenter gave the participants a general explanation of the research purpose and experimental procedure. After the explanation, informed consent was obtained from the participants. The experimental procedure consisted of the following three sessions: (a) the reading session (i.e., the standard condition followed by the task condition), (b) the English-reading proficiency test, and (c) the cued recall task.

Reading session

Figure 4.6 illustrates the procedure of Experiment 4. At the beginning of the reading session, each participant was given one of the four booklets. The whole texts were presented at a time on one page of the booklet. The participants were told to read the texts for comprehension at their own pace to answer the yes–no comprehension questions after reading (i.e., the standard condition). They were also told to verbalize any thoughts that came to their minds during reading. The following instructions were given to the participants:

> Please verbalize in Japanese any thoughts that come to mind while reading each text. What you verbalize will not be scored as correct or incorrect. Please verbalize any

thoughts, even if they seem trivial or not directly related to the texts. You do not have to explain what you mean to the experimenter. Imagine that you are talking to yourself aloud.

After the participants thought that they understood each text sufficiently, they moved on to the next page of the booklet and answered the yes–no comprehension question about the texts they had read. When answering the questions, the participants were not allowed to refer to the text on the previous page.

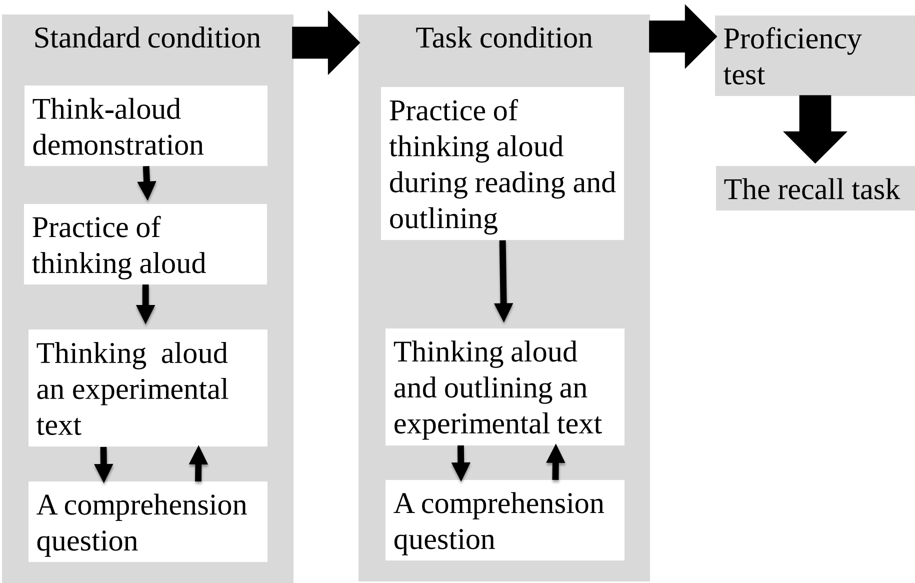

Figure 4.6. Procedure of Experiment 4.

Because the participants were not familiar with the think-aloud procedure, a practice session was conducted before reading the experimental texts. The researcher first gave the participants a practice text and had them listen to a recording of a think-aloud completed by an undergraduate student to demonstrate the process. After listening to the demonstration, the participants themselves practiced thinking aloud with another practice text and then answered a yes–no comprehension question. The experimenter intervened when (a) the participants read without saying anything for a while (e.g., “Please verbalize any thoughts that come to mind”), (b) they talked too quietly (e.g., “Please talk more loudly”), and (c) they only translated the practice text into Japanese or read aloud in English without expressing their thoughts (e.g., “Although you are allowed to translate the text into Japanese and read aloud in English, please verbalize any other thoughts you are having while reading”). The main purpose of these interventions was to familiarize the participants with verbalizing their thoughts about the contents of the experimental texts. These interventions were also provided when the participants were reading the experimental texts. After the practice session, the participants participated in the main experiment: reading two experimental texts for comprehension to answer the yes–no comprehension questions in the standard condition.

After reading in the standard condition, the participants read the other two experimental texts in the task condition. Before reading in the task condition, another practice session was conducted. The participants were first given the explanation of the task condition. The instructions were almost the same as the outline instructions given in Experiment 3, with the exception of the following two changes: (a) The participants were told to write the outlines while reading instead of simply reading to prepare for the outline task after reading, and (b) they were instructed to verbalize their thoughts while outlining the texts as well as while reading. To familiarize themselves with the procedure of the task condition, the participants completed a think aloud for a practice text while reading and writing the outline, and then they answered a yes–no comprehension question. At the end of the second practice session for the task condition, the participants were presented with an example of the outline of the practice text, as in Experiment 3.

After reading the two experimental texts in the task condition, the participants answered the English-reading proficiency test within 30 minutes. After finishing the proficiency test, the participants completed the written recall task, with the introductory sentence of each text presented as a recall cue. The time was not limited for the recall task, so that the participants had sufficient time to write down everything they remembered.

4.2.2.4 Scoring

Think-aloud protocols

The think-aloud protocols of each participant was transcribed and then parsed into clauses. Each clause was classified based on the type of reading process. To classify the think-aloud comments, this study adapted the frameworks of Horiba (2013) and Shimizu (2015). These studies examined EFL reading processes with theoretical focuses similar to that of the present study (i.e., the effect of educational interventions on reading processes, reading processes that build global coherence). Each clause was classified into one of the following categories: (a) analysis, (b) inference, (c) reader response, (d) rereading, (e) monitoring, (f) text structure, and (g) others. Descriptions of the categories are shown in Table 4.16.

The category of analysis represented surface-level understandings, analyzing the form or meaning of individual words, phrases, or sentences. This category included translating and paraphrasing the texts into Japanese. The category of inference included in-text inferences in which the participants attempted to link focal sentences with preceding sentences and to predict incoming information (i.e., backward and predictive inferences). The category of reader response included the participants' attempts to link the contents of the texts with their own knowledge. The category of reader response included think-aloud comments not directly related to the participants' understanding of the texts. However, the comments were still part of an active reading process that allowed the readers to relate to the text. This category included making associations, evaluating the contents of the texts, and reacting emotionally to text information. The category of rereading included the reiteration of preceding text information in the participants' minds. The category of monitoring included the participants' comments about the degree of their text comprehension or use of reading strategies. The category of text structure included the participants' references to text structure and their recognition of the role/importance of specific pieces of information in the texts. This category included comments about topic structure. To address RQ 4-1, the number of think-aloud protocols (which was equal to the number of the texts) was calculated, in which the hierarchical relations between the major topics and subtopics were identified.

Table 4.16

Categories and Descriptions of the Think-Aloud Comments

Category	Description
Analysis	The participants attempted to analyze the form or meaning of individual words and sentences, including L1 translation and paraphrasing.
Inference	The participants attempted to link the focal sentences with prior sentences or anticipate what information would appear next in the texts.
Reader response	The participants associated the contents of the texts with their own knowledge, made evaluative comments about the contents of the texts, or reacted emotionally to the texts.
Rereading	The participants reiterated what they had already read.
Monitoring	The participants commented on the degree of their text comprehension and use of reading strategies.
Text structure	The participants attempted to comment on text structure and the role/importance of specific pieces of information in the texts.
Others	The participants made comments unrelated to the texts or the outline task.

Two graduate students majoring in English education scored the think-aloud protocols collected from 30% of the participants. The kappa coefficient as inter-rater reliability was $\kappa = .62$, $p < .001$. All disagreements were resolved in discussion and one of the raters scored the rest of the comments. The frequency and proportion of each category were calculated.

Recall protocols

In Experiment 4, memory of topic structure was evaluated by the quality of the recall protocols. That is, the recall protocols were scored based on the degree to which they represented the topic structure of the texts. Table 4.17 displays an example of a recall protocol categorized as sufficiently representing the topic structure. To this end, Ghaith and Harkouss's (2003) criteria were adapted for the analysis, and Table 4.18 shows the scoring criteria. Two graduate students majoring in English education scored the recall protocols from 30% of the participants. The kappa coefficient was $\kappa = .79$, $p < .001$. All disagreements were resolved through discussion, and one of the raters scored the rest of the recall protocols.

Table 4.17

An Example of the Recall Protocols Scored as Sufficiently Representing the Topic Structure

Chimpanzees have unique lifestyles compared with other animals. <u>Chimpanzees live in a complex society.</u> Five or six males live with females and babies. While the females take care of the babies, the adult males protect the group. There is a hierarchy among the adult males. <u>Another characteristic of chimpanzees is that they communicate with hand signals</u>. The reason is that hand signals are silent. When chimpanzees find an enemy, hand signals enable them to inform each other of the enemy without making any noise or being discovered.

Note. The example of the recall protocol was translated into English by the author. The major topic is boldfaced and the subtopics are underlined.

Table 4.18

Criteria for the Qualitative Scoring of the Recall Protocols

Category	Criteria
Full recall	The major topic was produced, followed by at least one subtopic or one cluster consisting of the subtopic followed by its supporting details in the same paragraph.
Partial-major recall	The major topic was produced, but clusters of the subtopics followed by the supporting details were not produced. Specifically, the subtopics were not produced, or the subtopics and supporting details were produced in random order, like a list.
Partial-sub recall	The major topic was not produced. However, at least one subtopic or one cluster of the subtopic followed by the supporting details were produced.
Others	The major topic and subtopics were not produced or produced in random order, like a list.

Text outlines

In scoring the text outlines written in the task condition, two graduate students majoring in English education determined whether the participants produced correct text information as the major topics and subtopics. Specifically, (a) the outlines included the correct major topic and at least one correct subtopic (i.e., full outline), (b) the outlines included the correct major topic but not the correct subtopics (i.e., partial-major outline), (c) the outlines included at least one correct subtopic but not the correct major topic (i.e., partial-sub outline), and (d) the outlines did not include the correct major topic or subtopics (i.e., others). The kappa coefficient as the inter-rater reliability was $\kappa = .62$, $p < .001$. All rating discrepancies were resolved through discussion. Based on the discussed criteria, one of the raters scored the rest of the outlines.

4.2.2.5 Analysis

To examine the effect of writing outlines on general reading processes, a 2 (proficiency: higher, lower) × 2 (reading: standard, task) × 6 (category) three-way mixed ANOVA was carried out for the proportions of each category of the think-aloud comments. Proficiency was treated as a between-participant variable, whereas reading and category were treated as within-participant variables. To analyze the specific effect of the outline task on topic structure processing during reading, the number of the think-aloud comments (i.e., the number of the texts) in which the topic structure was identified was compared between the standard and task conditions (H4-1 and RQ4-1). To this end, two 2 (reading) × 2 (topic structure processing: successful, unsuccessful) Fisher's exact probability tests were carried out for the higher- and lower-proficiency groups. In addition, to explore the differences between the proficiency groups, a 2 (proficiency) × 2 (topic structure processing) Fisher's exact probability test was performed for the task condition. (In the standard condition, no participants identified topic structure in their think-aloud protocols.) Because three statistical tests were conducted for the think-aloud comments that identified topic structure, the alpha level was set at $\alpha = .017$, adopting the Bonferroni correction. The higher- and lower-proficiency groups were also compared in terms of the number of outlines. A 2 (proficiency) × 4 (category: full, partial-major, partial-sub, others) Fisher's exact probability test was conducted for the comparison.

As for the recall protocols, to examine the effect of writing outlines on reader memory of topic structure (the H4-2 and RQ4-2), two 2 (reading) × 4 (category) Fisher's exact probability tests were conducted for the number of recall protocols collected from the higher-and lower-proficiency groups. In addition, to examine differences according to L2 reading proficiency, two 2 (proficiency) × 4 (category) Fisher's exact probability tests were performed for the number of the recall protocols in the standard and task conditions. Because four statistical tests were conducted for the recall protocols, the alpha level was set at $\alpha = .13$, adopting the Bonferroni correction. Because there were cells in which the expected values were less than five, Fisher's exact probability tests were used instead of chi-squared tests.

Furthermore, the relations between understandings of topic structure while writing outlines and reader memory of it were analyzed for each proficiency group in the task condition. Specifically, Spearman's rank correlation coefficients were calculated between the number of outlines and the recall protocols that identified topic structure.

4.2.3 Results

4.2.3.1 English-Reading Proficiency Test

Table 4.19 shows the descriptive statistics of the English-reading proficiency test (Cronbach's $\alpha = .87$). The participants were divided into higher- and lower-proficiency groups based on the median of the test scores. A *t* test confirmed that the higher-proficiency group surpassed the lower-proficiency group in test scores, $t(30) = 7.37$, $p < .001$, $d = 2.61$.

The English-reading proficiency of the higher-proficiency group was approximately between the second and pre-first grade levels of the EIKEN test (levels B1 and B2 on the CEFR; Dunlea, n.d.). Their correct answer rates were over 60% for the second grade test but below 70% for the pre-first grade test. The lower-proficiency group was estimated to have an English-reading proficiency below the second grade (level B1 level on the CEFR; Dunlea, n.d.). Their correct answer rates were below 70% and 60% for the pre-first and second grade tests, respectively.

Table 4.19

Descriptive Statistics of the English-Reading Proficiency Test in Experiment 4

		Second (k = 20)		Pre-first (k = 8)		Total (k = 28)	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Higher	16	16.50	2.00	1.81	2.40	18.31	3.66
Lower	16	9.81	2.71	0.25	0.77	10.06	2.57

4.2.3.2 Yes–No Comprehension Questions

Table 4.20 shows the descriptive statistics of the yes–no comprehension questions. The correct answer rates were higher than 90% regardless of the proficiency group or reading condition. From this result, it was confirmed that the participants did not have any difficulty with text comprehension.

Table 4.20

Descriptive Statistics of the Yes–No Comprehension Questions in Experiment 4

		Standard			Instruction		
	n	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Higher	16	1.00	n/a	0.00	.94	[.85, 1.03]	.17
Lower	16	1.00	n/a	0.00	.97	[.90, 1.04]	.13

Note. 95% CIs were unavailable for the standard condition because all answers were correct.

4.2.3.3 Think-Aloud Protocols

Table 4.21 and Figure 4.7 show the descriptive statistics of the proportions of each think-aloud category. To examine the effect of the outline task on reading processes, a 2 (proficiency: higher, lower) × 2 (reading: standard, task) × 6 (category) three-way mixed ANOVA was carried out for the proportion of each category. The result revealed that a main effect of proficiency, $F(1, 30) = 7.79$, $p = .009$, $\eta_p^2 = .21$, and a main effect of category, $F(2.22, 66.45) = 479.24$, $p < .001$, $\eta_p^2 = .94$, were significant. Additionally, a main effect of reading approached significance, $F(1, 30) = 3.44$, $p = .074$, $\eta_p^2 = .10$. However, more importantly, a Reading × Category interaction was significant, $F(2.29, 68.57) = 39.09$, $p < .001$, $\eta_p^2 = .57$, qualifying the main effects of reading and category. This interaction suggested that the outline task affected reading processes in different ways, depending on the think-aloud category. None of the other interactions were significant (all *p*s > .10).

Table 4.21

Descriptive Statistics of the Proportions of Each Think-Aloud Category

	Higher-proficiency (*n* = 16)				Lower-proficiency (*n* = 16)			
	Standard		Task		Standard		Task	
Category	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Analysis	.81	.13	.60	.13	.81	.14	.60	.13
Inference	.04	.04	.02	.03	.03	.03	.03	.02
Response	.05	.06	.03	.04	.04	.04	.05	.05
Rereading	.06	.08	.14	.09	.04	.07	.12	.09
Monitoring	.03	.05	.05	.06	.06	.06	.08	.05
Structure	.02	.03	.15	.11	.02	.04	.12	.09

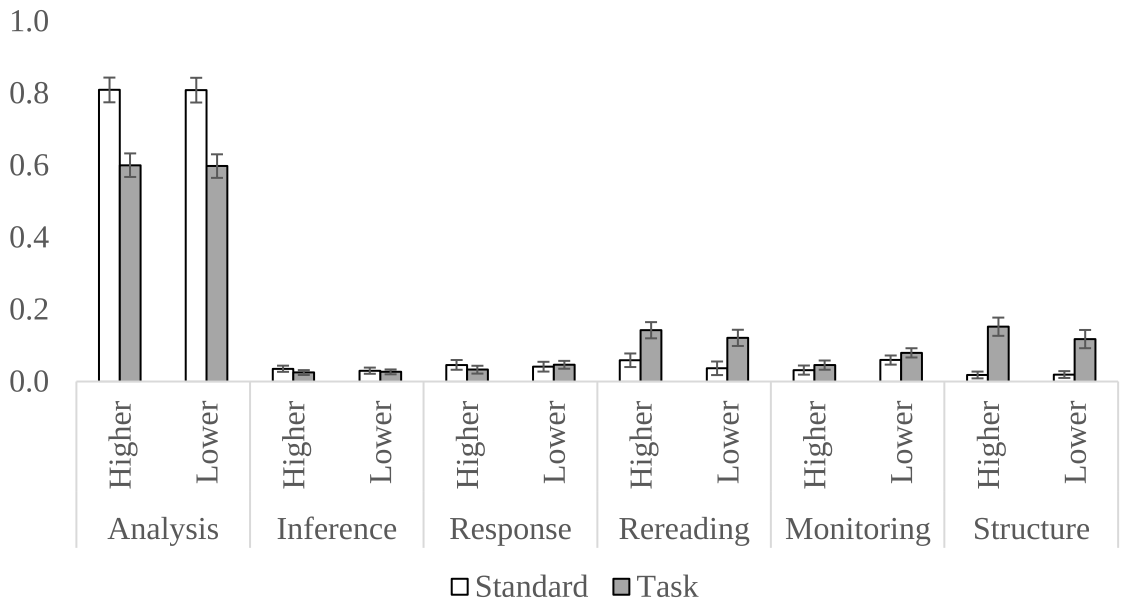

Figure 4.7. Mean proportions of each think-aloud category (± *SEM* bars).

To specify the task effect on each think-aloud category, a simple main effect of reading was tested for each category, showing that it was significant for analysis, rereading, and text structure (all ps < .001). Specifically, the participants decreased the proportion of analysis from the standard to the task condition. Although the participants allocated more than 80% of their cognitive resources to surface level understandings in the standard condition, they decreased their resource allocation to word- and sentence-level processes when engaging in the outline task. On the other hand, they spent more cognitive resources on rereading to write the outlines, indicating that they attempted to read more carefully and review the previous sentences. Moreover, the outline task made the participants pay more attention to text structure. Table 4.22 summarizes the results of the three-way ANOVA.

Table 4.22

Summary of the Three-Way ANOVA for the Proportions of Each Think-Aloud Category

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
			Between-participants			
Proficiency (P)	< 0.01	1.00	< 0.01	7.79	.009	.21
Error (P)	< 0.01	30.00	< 0.01			
			Within-participants			
Reading (R)	< 0.01	1.00	< 0.01	3.44	.074	.10
R × P	< 0.01	1.00	< 0.01	1.22	.279	.04
Error (R)	< 0.01	30.00	< 0.01			
Category (C)	22.40	2.22	10.11	479.24	< .001	.94
C × P	0.03	2.22	0.01	0.59	.574	.02
Error (C)	1.40	66.45	0.02			
R × C	1.04	2.29	0.45	39.09	< .001	.57
R × C × P	0.01	2.29	< 0.01	0.26	.803	.01
Error (R × C)	0.80	68.57	0.01			

Table 4.23 shows the number of think-aloud protocols (i.e., the number of texts) in which topic structure was identified. There were no think-aloud protocols identifying topic structure in the standard condition, regardless of the proficiency group, confirming the difficulty of independent topic structure processing for EFL readers. To reveal whether the outline task facilitated topic structure processing, 2 (reading) × 2 (topic structure processing) Fisher's exact probability tests were performed for the higher- and lower-proficiency groups (the alpha level was corrected to $\alpha = .017$). Whereas the number of think-aloud protocols significantly increased from the standard to the task condition in the higher-proficiency group

(p < .001), this positive effect of the outline task did not appear for the lower-proficiency group (p = .053). Moreover, to explore the difference according to L2 reading proficiency, a 2 (proficiency) × 2 (topic structure processing) Fisher's exact probability test was carried out for the number of think-aloud protocols in the task condition. The result did not yield a significant difference (p = .088). Combining the above results, the outline task seemed to help the higher-proficiency group understand topic structure during reading but did not help the lower-proficiency group, although a significant difference related to L2 reading proficiency did not appear in the task condition.

Table 4.23

The Number of Think-Aloud Protocols That Identified Topic Structure

Higher-proficiency				Lower-proficiency			
Standard		Task		Standard		Task	
n	%	*n*	%	*n*	%	*n*	%
n/a	n/a	12	37.50	n/a	n/a	5	15.63

Note. Thirty-two think-aloud protocols were collected as a function of the proficiency groups and reading conditions. The higher- and lower-proficiency groups (n = 16 in each) completed the think-aloud process for two texts in each reading condition.

4.2.3.4 Outline Task

Table 4.24 and Figure 4.8 show the number of the outlines classified into each category. To compare the two groups' performances on the outline task, a 2 (proficiency: higher, lower) × 4 (category: full, partial-major, partial-sub, others) Fisher's exact probability test was conducted for the number of the outlines. The result showed different tendencies for each group although the difference was not significant (p = .073). Residual analysis as a post hoc test indicated that the higher-proficiency group produced more full and fewer partial-sub outlines than the lower-proficiency group (adjusted residuals = 2.18, –2.00, respectively). The outline performances suggest that the higher-proficiency group tended to understand topic structure better than the lower-proficiency group. On the other hand, although the lower-proficiency group understood clusters of the subtopics and supporting details, this group tended to be worse at linking the clusters with the major topics than the higher-proficiency group.

Table 4.24

The Number of Outlines Classified Into Each Category

	Higher-proficiency		Lower-proficiency	
Category	*n*	%	*n*	%
Full	18	56.25	9	29.03
Partial-major	4	12.50	2	6.45
Partial-sub	6	18.75	13	41.94
Others	4	12.50	7	22.58

Note. The higher- and lower-proficiency groups (n = 16 in each) wrote outlines for two texts in the task condition.

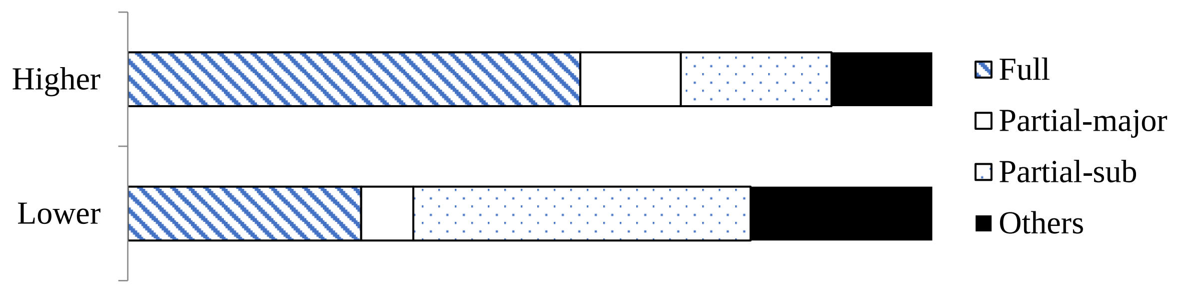

Figure 4.8. Proportions of the outlines classified into each category.

4.2.3.5 Recall Task

Table 4.25 and Figure 4.9 show the number of the recall protocols classified into each category. The recall quality was compared according to L2 reading proficiency and reading condition, and four statistical tests were conducted (the alpha level was corrected to $\alpha = .013$). First, the recall quality in the standard condition was analyzed according to reading proficiency. A 2 (proficiency: higher, lower) × 4 (quality: full, partial-major, partial-sub, others) Fisher's exact probability test was performed. This analysis did not show a significant difference between the proficiency groups ($p = .835$). In the standard condition, most of the recall protocols were scored as belonging to the partial-sub or other category, regardless of reading proficiency level. The participants' text memory was poorly organized or only represented paragraph-level clusters consisting of the subtopics and supporting details. Additionally, the same Fisher's exact probability test was carried out for the task condition. This result suggests that the recall quality tended to differ according to L2 reading proficiency ($p = .015$), although this difference was not found to be significant. Post hoc tests showed fewer recall protocols were categorized as partial-sub for the higher- than the lower-proficiency group (adjusted residual = –2.89).

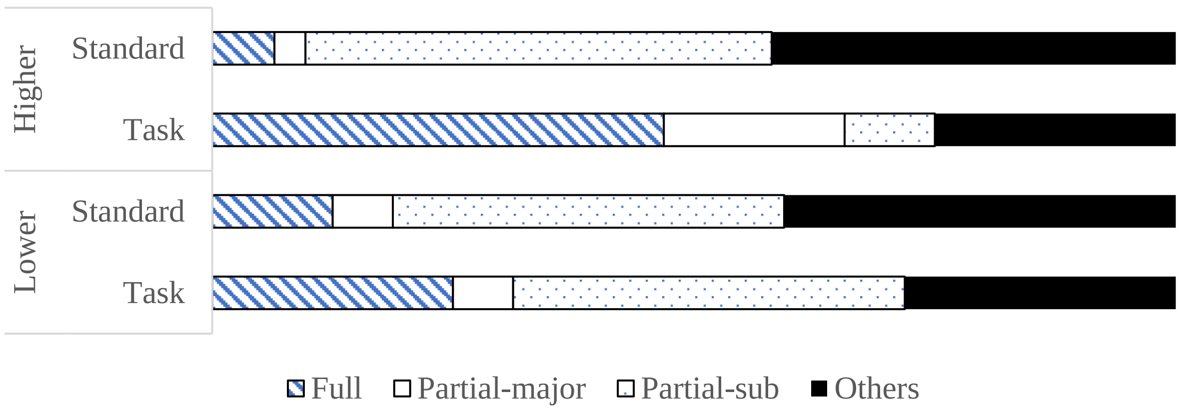

Figure 4.9. Proportions of the recall protocols classified into each category.

Table 4.25

The Number of Recall Protocols Classified Into Each Category

	Higher-proficiency				Lower-proficiency			
	Standard		Task		Standard		Task	
Category	*n*	%	*n*	%	*n*	%	*n*	%
Full	2	6.45	15	46.88	4	12.50	8	25.00
Partial-major	1	3.23	6	18.75	2	6.25	2	6.25
Partial-sub	15	48.39	3	9.38	13	40.63	13	40.63
Others	13	41.94	8	25.00	13	40.63	9	28.13

Note. The higher- and lower-proficiency groups (n = 16 in each) completed the recall task for two texts in each reading condition.

The recall quality was also compared among the reading conditions. A 2 (reading: standard, instruction) × 4 (quality: full, partial-major, partial-sub, others) Fisher's exact probability test was conducted for each proficiency group. The result demonstrated that the recall quality of the higher-proficiency group differed among the reading conditions ($p < .001$), while differences among the reading conditions were not observed for the lower-proficiency group ($p = .537$). Post hoc tests revealed that the full and partial-major categories increased from the standard to the task condition (adjusted residuals = 3.61, 1.96), whereas the partial-sub category decreased (adjusted residual = –3.43). This result indicates that the outline task helped the higher-proficiency group with the understanding of overall topic structure, instead of simply paragraph-level understandings. On the other hand, the outline task failed to affect text memory for the lower-proficiency group.

4.2.3.6 Relation Between the Outline and Recall Performance

To investigate the relations between understandings of the topic structure during writing outlines and memory of it, Spearman's rank correlation coefficients were calculated between the number of the outlines and recall protocols that represented topic structure (i.e., the full category) in the task condition. The result revealed that there was a significant and positive correlation in the higher-proficiency group (r = .61, p = .013) and in the lower-proficiency group (r = .82, p < .001). This suggests that writing outlines while reading contributes to reader understanding and memory of topic structure.

4.2.4 Discussion

Difficulty understanding topic structure without engaging in a specific task (Hs4-1 and 4-2)

The participants uttered no think-aloud comments that identified topic structure in the standard condition, which supported H4-1. As past L2 research has indicated, cognitive resources available for higher-level processes such as topic structure processing were probably competing with lower-level processes (e.g., word recognition, syntactic parsing), which are not proficient in L2 readers compared to L1 readers (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016; Yoshida, 2003). This was made evident by the proportion of the think-aloud comments in the standard condition. Specifically, analysis exceeded 80%, whereas other categories were below 10%, regardless of L2 reading proficiency. As the minimalist hypothesis (McKoon & Ratcliff, 1992) assumes, the participants could not link the major topics with subtopics without using a strategic process. In addition, the participants might have set their standards of coherence at the sentence level and did not attempt to build global coherence (e.g., van den Broek et al., 2015); this is likely because they were simply told to read the texts to answer the yes–no comprehension questions asking about the supporting details.

The readers' difficulty with linking subtopics with major topics during reading for comprehension was consistent with the findings of Experiments 2A, 2B, and 3. In these experiments, reader understanding of topic structure might have been made more difficult because the texts were presented sentence by sentence. That is, the participants in these experiments were not allowed to look back at previous sentences and reactivate the major topics and subtopics, in contrast to other previous studies (e.g., Hyönä & Lorch, 2004; Hyönä et al., 2002; Hyönä & Nurminen, 2006). However, this possibility can be rejected by comparing the standard conditions of Experiments 3 and 4. Even though the participants read the entire texts in Experiment 4, they were not engaged in topic structure processing. This comparison shows that topic structure processing is difficult for L2 readers regardless of the manner of text presentation. Thus, other factors, such as limited cognitive resources during EFL reading, appear to be more relevant to the readers' failure to understand topic structure.

As for the readers' recall ability, few of the recall comments were categorized as full recall in the standard condition (i.e., 6.45% and 12.50% for the higher- and lower-proficiency groups, respectively), which supports H4-2. Previous studies have shown that readers comprehend topic structure during reading and can use that structure to retrieve text memory in post-reading tasks (e.g., Britton, 1994; Hyönä & Lorch, 2004; Lorch & Lorch, 1985; Lorch et al., 2001). However, because the participants of the present study were unsuccessful in topic structure processing while reading, they were not able to use topic structure to retrieve text memory. Furthermore, the results of the present study suggest that the participants were not able to understand topic structure through the reconstruction of text memory in the recall task.

This difficulty with globally coherent comprehension left the participants' text memory fragmented. In fact, most of the recall protocols in the standard condition were classified into the others (i.e., 41.94% and 40.63% for the higher- and lower-proficiency groups, respectively) and partial-sub categories (i.e., 48.39% and 40.63%). Hence, the readers' text memories were unorganized, or they were only able to represent links between the subtopics and supporting details within a paragraph. This view is consistent with Kimura's (2014) finding that EFL readers who failed to understand the overall themes of expository texts tended to mistakenly identify the paragraph subtopics as themes.

The effect of writing outlines on EFL reading processes (RQ 4-1)

The proportions of the think-aloud comments showed that the participants changed their reading processes from the standard to the task condition, regardless of L2 reading proficiency. Specifically, the participants decreased resource allocation to analysis and increased rereading and comments to text structure. By increasing rereading, the participants read the texts more attentively and attempted to reactivate prior sentences in their minds. Additionally, they paid more attention to text structure in an attempt to construct globally coherent comprehension. Although increases in rereading and comments on text structure were observed in both proficiency groups, the specific contents of these reading processes differed according to L2 reading proficiency.

As for rereading, the higher-proficiency group selectively looked back at the major topics and subtopics, whereas the lower-proficiency group reread more frequently but non-selectively. Among the think-aloud comments categorized as rereading in the task condition, more rereading was observed for the major topics and subtopics in the higher-proficiency group (37/88 [42%] comments) than in the lower-proficiency group (28% [31/112] comments). A similar tendency has also been observed in L1 reading processes (Hyönä et al., 2002; Hyönä & Nurminen, 2006). Because it is necessary to reactivate distant prior sentences to build globally coherent comprehension (e.g., van der Schoot et al., 2012), the participants may have attempted to reactivate the major topics and subtopics by rereading.

Regarding the comments on text structure, writing outlines improved reader understanding of topic structure in the higher-proficiency group (12/32 [38%] texts in the task condition), but this was not effective for the lower-proficiency group (5/32 [16%] texts in the task condition). Although reader understanding of topic structure did not differ by L2 reading proficiency in the think-aloud protocols, the difference was apparent in the participants' performance on the outline task. Specifically, the lower-proficiency group produced more partial-sub outlines, only maintaining coherence within paragraphs. On the other hand, the higher-proficiency group tended to write fuller outlines that better represented the topic structure of the texts.

However, it should be noted that all the think-aloud comments that identified topic structure were reported after the participants read the entire texts. That is, they did not engage in topic structure processing or write outlines in the middle of reading; rather, they wrote outlines and understood the topic structure after reading the texts. Similarly, Ponce and Mayer (2014) suggested that L1 readers made graphic organizers after reading the whole texts as a strategic learning strategy (i.e., selecting and organizing text information to make graphic organizers). Because writing outlines also required the selection and organization of text information, the participants might have strategically written the outlines after reading all the entire texts.

As shown above, writing outlines had different effects on the higher- and lower-proficiency groups, which was explained by two possible theories. The positive effect of writing outlines on the higher-proficiency group was supported by activity theory, which hypothesizes that deep comprehension, such as selection and organization of text information, is achieved by engaging in a productive task (e.g., Stull & Mayer, 2007). In the present study, writing outlines helped the higher-proficiency group to select the major topics and subtopics to be included in the outlines and then organized them in accordance with the hierarchical relations in the topic structure. On the other hand, the ineffectiveness of writing outlines for the lower-proficiency group can likely be explained by cognitive load theory (e.g., Sweller, 1988). This theory suggests the possibility that engagement in productive tasks demands too much mental effort, which mitigates its positive effects. Thus, writing outlines might have been too cognitively demanding for the lower-proficiency group because the task required them to select and organize the major topics and subtopics simultaneously (Lorch et al., 2013).

Possible reasons for the differences according to L2 reading proficiency were inspected in further detail. Although Ushiro et al. (2009) suggested that L2 reading proficiency affected text comprehension rather than globally coherent comprehension, both proficiency groups in this study sufficiently understood the texts. (The correct answer rates of the comprehension questions were 94% and 97% for the higher- and lower-proficiency groups, respectively.) Instead, in relation to cognitive load theory as mentioned above, it is possible that the higher-proficiency group was relatively more proficient in lower-level processes than the lower-proficiency group and thus had more cognitive resources available for the outline task. This is consistent with the findings of previous studies that showed that readers with more cognitive resources succeeded more often in understanding topic structure (Hyönä et al., 2002; Hyönä & Nurminen, 2006). Furthermore, another possibility is that L2 reading proficiency might not have influenced paragraph-level comprehension in past research (Ushiro et al., 2009) but affected linking information across paragraphs in the present study. This is consistent with the findings of Kimura (2014), who found a similar effect of L2 reading proficiency on the thematic comprehension of expository texts.

The effect of writing outlines on text memory after EFL reading (RQ4-2)

The results of the recall task demonstrated that the higher-proficiency group reduced partial-sub recalls and increased full recalls from the standard to the task condition. This is consistent with the result of the think-aloud comments and outlines that showed the positive effect of writing outlines on the participants' understanding of the topic structure. Moreover, the significant correlation between reader understanding of topic structure in the outline and recall performance suggests that the higher-proficiency participants represented topic structure in their text memory through better understanding during reading and outlining the texts. Previous studies have found that when readers understand the topic structure of texts, they can use it to assist in the retrieval of text memory in post-reading tasks (e.g., Hyönä & Lorch, 2004; Lorch & Lorch, 1985; Lorch et al., 2001). Specifically, the retrieval of a major topic provides access to its subtopics in text memory.

Moreover, as much as the understanding of topic structure itself contributed to text recall, the cognitive processes used to complete the outline task might have occurred in the reconstruction of text memory in the post-reading task. Specifically, outlining processes, such as identifying the major topics and subtopics and organizing them in a hierarchical manner, might have contributed to retrieving the readers' memory of topic structure. This view is possible because even EFL readers can afford to allocate cognitive resources to higher-level processes in post-reading tasks after finishing the resource-demanding, lower-level processes (Hosoda, 2014; Nahatame, 2013; Ushiro, Mori, et al., 2016).

In contrast to the higher-proficiency group, most recall protocols from the lower-proficiency group were classified into the categories of partial-sub recalls or others, regardless of the reading condition. This is consistent with Kimura's (2014) finding that Japanese university students tended to answer with the paragraph subtopics when they were not able to comprehend the overall themes of expository texts in a post-reading task. In other words, the readers were not able to integrate text information beyond the paragraph level, even when they attempted to do so. The ineffectiveness of outlining on topic structure memory is consistent with the results of the think-aloud comments and outlines. The relation between poor outlining and topic structure memory is also suggested by the significant correlation. Specifically, although the lower-proficiency group changed some of their cognitive processes in thinking aloud, they could not comprehend topic structure and write hierarchical outlines, which caused text memory to be fragmented or clustered into paragraphs. As discussed in RQ4-1, it is likely that cognitive resources were mostly allocated to lower-level processes (e.g., Morishima, 2013) and their ability to link information beyond paragraphs was poor (Kimura, 2014).

However, it should be also noted that the correlation between the outlines and recalls could be interpreted in another way. That is, even the lower-proficiency participants likely represented topic structure in text memory when comprehending it in the outline task, although they had difficulty outlining the texts. Further, combining the proficiency effect in the outlines and recalls with the correlation observed in both proficiency groups, L2 reading proficiency mainly affected outline performance. In addition, once the participants understood topic structure in the outline task, their understanding was represented in text memory, regardless of proficiency level.

4.2.5 Conclusion of Experiment 4

The purpose of Experiment 4 was to reveal whether writing outlines during reading helped Japanese EFL readers understand topic structure during reading and represent it in their text memory. The main findings of Experiment 4 are summarized below in four points.

First, the participants were not able to understand topic structure during reading for comprehension, which supports H4-1. It was suggested that they failed to do so because most of their cognitive resources were consumed by lower-level processes, as discussed in the previous experiments of this study. Additionally, although Experiment 3 suggested the possibility that the sentence-by-sentence reading method prohibited the participants from adequately processing topic structure, the participants failed at that processing even when the entire texts were presented in Experiment 4. Thus, the participants seemed to have difficulty understanding topic structure, regardless of how the texts were presented.

Second, the participants did not represent topic structure in their text memory after reading for comprehension, which supports H4-2. This is probably due to the participants' failure to process topic structure while reading. Their memory of most texts was fragmented or clustered into paragraphs. This result is consistent with the findings of previous studies (Kimura, 2014) that showed that EFL readers tend to understand the subtopics of paragraphs while failing to understand the overall themes of expository texts.

Third, writing outlines helped the higher-proficiency group to better understand topic structure and selectively reread the texts to reactivate the major topics and subtopics (RQ4-1). However, it should be noted that topic structure was identified during writing outlines after reading the whole texts. This might represent their strategy to write outlines, and select and organize information from entire texts. On the other hand, the lower-proficiency group still struggled to understand topic structure and reread sentences that were irrelevant to topic structure. It is possible that the outline task was too resource demanding for the lower-proficiency group, or they had insufficient linking skills for globally coherent comprehension beyond paragraphs.

Finally, writing outlines improved the memory of topic structure in the higher-proficiency group, but it was not effective for the lower-proficiency group (RQ4-2). The higher-proficiency group displayed better comprehension of topic structure on the outline task, and that comprehension might have served as a retrieval method for text recall. In addition, it is possible that selection and organization of the major topics and subtopics in writing outlines supported the participants' reconstruction of text memory, which possibly contributed to their memory of topic structure. On the other hand, the text memory of the lower-proficiency group remained fragmented or clustered into paragraphs even after writing the outlines.

4.3 Conclusion of Study 2

Study 2 investigated whether and how educational interventions supported topic structure processing during EFL reading and its memory. Because the outline task required the participants to select and organize the major topics and subtopics, the effects of the outline instructions (Experiment 3) and engagement in the outline task (Experiment 4) were explored.

Experiment 3 examined the effect of the outline instructions on reader understanding of topic structure during reading and its memory, by comparing the reading times and recall rates of the subtopics between the texts with and without explicit major topics. When the outline instructions were given, the participants read the texts more attentively and their text memory was enhanced. However, the outline instructions failed to support reader understanding of topic structure during reading and the recall task. One possible explanation is that the outline instructions did not sufficiently affect text comprehension because the participants did not actually do the outline task.

Hence, Experiment 4 investigated the effect of writing outlines on the readers' understanding of topic structure during reading and its memory, adopting the think-aloud method and analyzing recall quality. The outline task helped the higher-proficiency group to better understand topic structure and reactivate relevant information in their text memory (i.e., the major topics and subtopics) while writing outlines after reading the texts. Moreover, this enhanced memory of topic structure and reconstruction of text memory in the recall task. On the other hand, although the lower-proficiency group changed their reading processes, these changes were not sufficient for better understanding or memory of topic structure. The lower-proficiency group might not have been proficient in the required lower-level processes, which limited the cognitive resources available for the outline-writing task; moreover, they may have had poor linking skills beyond paragraphs.

In summary, Study 2 revealed the effects of educational interventions on L2 readers with different reading proficiency levels. For the higher-proficiency readers, although giving the outline instructions made them more attentive and enhanced their text memory, it did not significantly improve their understanding of topic structure. However, they better understood topic structure while writing outlines after reading and represented it in their text memory. On the other hand, for the lower-proficiency group, neither the outline instructions nor the outline task improved their understanding or memory of topic structure. A possible explanation is that their lower L2 reading proficiency might have limited the available cognitive resources and skills to link information beyond paragraphs.

Chapter 5
General Discussion and Conclusion

5.1 General Discussion

5.1.1 Overview of Findings

To examine topic structure processing during EFL reading and reader memory of topic structure, the present study conducted a total of five experiments, addressing a total of four Hs and seven RQs. First, Study 1 conducted three experiments (i.e., Experiments 1 to 2B) to investigate whether Japanese EFL readers understood the links between the major topics and subtopics of expository texts, as well as the links between the subtopics and supporting details. The following three RQs were addressed in Study 1:

RQ1: Can Japanese EFL readers represent topic structure in their text memory?

RQ2-1: Can Japanese EFL readers process topic structure during reading that is measured by a recognition task?

RQ2-2: Can Japanese EFL readers process topic structure during reading that is measured by a lexical decision task?

Experiment 1 examined whether the Japanese EFL readers could link the subtopics with the major topics (RQ1) and supporting details in their text memory. The participants answered the immediate and delayed recall tasks when (a) a major topic, (b) a supporting detail, or (c) no information was presented as a recall cue. Comparison of the subtopic recall rates among the three cue conditions demonstrated that they linked the subtopics with the major topics and supporting details in their text memory, regardless of L2 reading proficiency and recall time.

Additionally, the subtopics were linked with the major topics more robustly than the supporting details in their text memory.

Experiment 2A examined whether the Japanese EFL readers linked the subtopics with the major topics (RQ2-1) and supporting details during reading. After reading the texts sentence by sentence on a computer screen, the participants answered the recognition task. In this task, the priming stimuli (i.e., a major topic, a supporting detail, or topically related but unsuggested information) were presented, and then the participants responded to the target words. Whereas the correct response times indicated that the links between the subtopics and major topics/supporting details were not readily available or activated in their minds, the correct response rates suggested that they recognized these links. Regarding the links between the subtopics and supporting details, these conflicting results suggest that the participants might have understood but deactivated the links during reading to ensure sufficient cognitive resources for incoming information. As for the links between the major topics and subtopics, the participants may have attempted to understand during reading but this resulted in poor understanding, or the recognition task led them to understand the links after reading.

Experiment 2B replicated Experiment 2A, with the exception that it used the lexical decision task instead of the recognition task (RQ2-2). Compared to Experiment 2A, the results were similar regarding the links between the subtopics and supporting details but different in terms of the links between the major topics and subtopics. That is, reader understanding of the links between the major topics and subtopics was indicated neither by the correct response times nor by the rates. Combined with Experiment 2A, the difficulty of topic structure processing during EFL reading and the possible effect of post-reading tasks were indicated by the results.

Study 2 conducted two experiments (i.e., Experiments 3 and 4) to examine the effect of educational interventions on reader understanding of topic structure, addressing the following four Hs and RQs:

H3-1: Japanese EFL readers have difficulty understanding topic structure during reading sentence by sentence.

H3-2: Japanese EFL readers can represent topic structure in their text memory.

H 4-1: Japanese EFL readers have difficulty understanding topic structure during reading when entire texts are presented at once.

H 4-2: Japanese EFL readers have difficulty representing topic structure in their text memory independently.

RQ 3-1: Do the outline instructions support topic structure processing during EFL

RQ 3-2: reading?

RQ 4-1: Do the outline instructions support memory of topic structure after EFL reading?

Does the outline-writing task support topic structure processing during EFL

RQ 4-2: reading?

Does the outline-writing task support memory of topic structure after EFL reading?

Experiment 3 examined whether the outline instructions helped the Japanese EFL readers to understand topic structure during reading and represent it in their text memory (RQ3-1, RQ3-2), compared to reading without specific reading instructions (H3-1, H3-2). The participants engaged in sentence-by-sentence reading and then completed the recall task. This experiment manipulated the explicitness of the major topics and compared the reading times and recall rates of the subtopics according to major-topic explicitness. The reading times and recall rates of the subtopics indicated that the participants failed to understand topic structure during reading or in the recall task without specific reading instructions. Regarding the effect of the outline instructions, although it encouraged the participants to read more attentively and enhanced their text memory, it failed to improve their topic structure processing during reading and memory of topic structure. Reader understanding of topic structure during reading and the recall tasks were found to be only weakly correlated.

Experiment 4 examined whether writing outlines supported topic structure processing during EFL reading and reader memory of it (RQ4-1, RQ4-2), compared to reading without engaging in a specific task (H4-1, H4-2). The participants engaged in a think-aloud process while reading the texts and then answered the recall task. When they did not engage in a specific task (i.e., control condition), the participants' think-aloud protocols did not identify the hierarchical relations between the major topics and subtopics, confirming their difficulty with topic structure processing while reading. Moreover, most of the recall protocols did not reflect understanding of the topic structure hierarchy.

When they wrote outlines during reading, the think-aloud comments demonstrated that the higher-proficiency group selectively reread the major topics and subtopics of the texts and better understood the topic structure, whereas the lower-proficiency group reread non-selectively and continued to have difficulty understanding topic structure. However, it should be noted that the higher-proficiency participants understood topic structure when they wrote the outlines after reading (i.e., not during the reading process). As for participant performance on the outline task, the higher-proficiency group wrote more outlines representing the overall topic structure and fewer outlines based on paragraphs, in comparison to the lower-proficiency group. In the recall task after writing outlines, the higher-proficiency group increased the recall protocols representing topic structure and reduced recall protocols based on paragraphs, whereas writing outlines failed to affect the recall quality of the lower-proficiency group. The outline and recall performance was significantly correlated in both the higher- and lower-proficiency groups. In summary, writing outlines helped the higher-proficiency group to understand the topic structure after reading and to represent it in their text memory. However, these tasks were not effective for the lower-proficiency group.

Based on the above summary of Studies 1 and 2, the following points will be discussed in the next subsection: (a) topic structure processing during EFL reading, (b) representations of topic structure in text memory, and (c) the effects of educational interventions on reader understanding of topic structure.

5.1.2 Topic Structure Processing During EFL Reading

Several theoretical frameworks can be used to explore how readers understand topic structure during reading. Specifically, Gernsbacher's (1990) structure building framework and Britton's (1994) text comprehension model hypothesize that readers use the major topics as contexts to understand the subsequent subtopics, locating the subtopics under the major topics to fit them into the hierarchy of topic structure. In addition, Kintsch's text comprehension model states that readers recursively integrate subordinate propositions into superordinate propositions (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983; Kintsch, 1998). In that process, readers are assumed to link the major topics with the subtopics. The above theoretical frameworks have been demonstrated in empirical research on L1 reading processes (e.g., Hyönä & Lorch, 2004; Murray & McGlone, 1997). However, the results of Experiments 2A to 4 revealed that Japanese EFL readers showed a different tendency from the theoretical assumptions of these frameworks, demonstrating that they had difficulty with topic structure processing during reading.

The results of Experiments 2A and 2B were inconsistent. In Experiment 2A, the correct response rates of the recognition task indicated reader understanding of topic structure, while the correct response times did not. On the other hand, in Experiment 2B, neither the correct response times nor rates of the lexical decision task indicated such understanding. There are two possible explanations for the difference between the experiments. First, because the participants were instructed to read for different priming tasks, they might have engaged in different during-reading processes. In Experiment 2A, the recognition task required the participants to refer to their text memory to determine whether the target words were explicitly mentioned in the texts (McKoon & Ratcliff, 1992). Hence, the participants might have attempted to construct more complete mental representations of the texts during reading to prepare for the recognition task.

However, this possibility can likely be rejected because the participants failed to understand topic structure when they were given more specific instructions. In Experiment 3, the participants were given the outline instructions. They were instructed to read the texts to prepare for itemizing and organizing the major topics and subtopics after reading. Although the outline instructions explicitly aimed to assist the participants in understanding the topic structure, they failed to do so. On the other hand, in Experiment 2A, the expectation of the recognition task was likely to affect during-reading processes in a less specific manner (i.e., understanding as many sentences as possible to recall later). Because the participants failed to adapt their during-reading processes to understand topic structure even when specific and direct instructions were given (i.e., reading for the outline task), they also did not engage in topic structure processing while reading when only unfocused instructions were given (i.e., reading for the recognition task).

Alternatively, another possibility to account for the difference between Experiments 2A and 2B is that the participants came to understand topic structure while answering the recognition task. As mentioned above, they had to refer to their text memory to answer the recognition task (McKoon & Ratcliff, 1984). In that process, the participants might have reconstructed their text memory, linking the major topics with subtopics. This might also explain why the correct response rates alone indicated reader understanding of topic structure in Experiment 2A. The present study adopted the recognition task and lexical decision task as on-line measurements. Jiang (2012) proposed that tasks can be treated as on-line measurements if they require quick responses and measure response times. In other words, whereas the correct response times seem to be more indicative of during-reading processes, the correct response rates seem less indicative of them, representing other cognitive processes that were occurring during the tasks. In the case of the present study, presenting the major topics increased the correct response rates in the recognition task but did not do so in the lexical decision task. This suggests that topic structure was understood by cognitive processes that occurred during the recognition task alone. That is, the participants referred to the text and then reconstructed their text memory. On the other hand, because the null effect of the major stimuli on the correct response times was observed regardless of the priming task, it indicates the general difficulty of topic structure processing during EFL reading.

The above possibility was further explored in the standard condition of Experiment 3. Although the correct response times in Experiments 2A and 2B suggest that the Japanese EFL readers had difficulty understanding topic structure while reading, these data were not exact on-line measurements that were collected immediately after reading. Jiang (2012) stated that some on-line measures are more or less indicative of during-reading processes than other on-line measures. Hence, Experiment 3 adopted a reading time method that measured real time processes during reading (Haberlandt, 1994). The results of the reading times more directly demonstrated the difficulty of topic structure processing during EFL reading, confirming that the second possibility discussed regarding Experiments 2A and 2B is more likely than the first possibility.

However, it is still possible that the difficulty with topic structure processing during EFL reading was due to a methodological factor rather than a more general phenomenon among EFL readers. That is, when the participants failed to understand topic structure in

Experiments 2A, 2B, and 3, they read the texts sentence by sentence. The sentence-by-sentence reading might have made it difficult to understand topic structure in the following two ways. First, the participants did not know where each sentence was located in the texts. As previous studies (Britton, 1994; Gernsbacher, 1990, 1997; Goldman et al., 1995) suggested, they might have used locational information (e.g., the first sentence of a paragraph is likely to be the topic of that paragraph) to look for the major topics and subtopics. However, this view was rejected by the importance rating task in Experiment 1, Experiment 2A, and the pilot study of Experiment 3. Although each sentence was presented in the lists and the locational information was unavailable, the different importance levels of the major topics and subtopics were identified. This suggests that the participants identified the major topics and subtopics based on their contents, even without using locational information. Indeed, previous studies have shown that readers used as many kinds of sources as possible to find important information (Goldman et al., 1995).

Second, the sentence-by-sentence reading did not allow the participants to reread the previous sentences. To build globally coherent comprehension, readers need to reactivate distant information from earlier parts of the texts (e.g., van der Schoot et al., 2012). In fact, L1 reading research has shown that L1 readers looked back at the major topics and subtopics during reading (e.g., Hyönä & Lorch, 2004; Hyönä et al., 2002; Hyönä & Nurminen, 2006), suggesting that they reactivated information relevant to the topic structure. Thus, the entire texts were presented in Experiment 4, and the results of the standard condition were compared to the results of Experiment 3. It was found that the participants failed to understand topic structure during reading, even when they were permitted to reread the preceding sentences. Because topic structure processing during reading was difficult regardless of the style of text presentation, there might have been common sources of difficulty for the participants across the experiments, such as the phenomenon of limited cognitive resources during EFL reading (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016; Yoshida, 2003).

As the above discussion of the present study's experiments indicates, the Japanese EFL readers had difficulty linking the major topics with subtopics while reading expository texts. This tendency was observed regardless of L2 reading proficiency. Such difficulties can be explained in terms of several theoretical frameworks. Regarding the automaticity of globally coherent comprehension during reading, the constructionist and minimalist hypotheses offer conflicting interpretations. The results of the present study are inconsistent with the assumptions of the constructionist hypothesis, which posits that readers build globally coherent comprehension without using specific reading strategies (Graesser et al., 1994). Rather, the findings of the present study are supported by the minimalist hypothesis, which argues that readers do not automatically construct globally coherent comprehension (McKoon & Ratcliff, 1992). In addition, the results of the present study are also consistent with the framework of the standards of coherence (van den Broek et al., 2015). This framework assumes that readers alter reading processes in response to their reading goals or the given tasks, which in turn affects text memory. In Experiment 2B and the standard condition of Experiments 3 and 4, the participants were told to read the texts for comprehension to answer the yes–no comprehension questions testing their literal comprehension of individual sentences. Hence, the participants were likely to set their standards of coherence at the sentence level, and simply attempted to comprehend the individual sentences literally (Ushiro et al., 2017). In Experiment 2A, the participants read the texts for the recognition task. Although the expectation of the recognition task might have motivated them to comprehend as many sentences as possible, this reading strategy was not likely to facilitate coherent, global text comprehension. Accordingly, the participants attempted to understand many sentences but could not link them, which led to fragmented text comprehension.

The difficulty with topic structure processing during reading also seemed consistent with the findings of past empirical studies. Some L1 studies have reported successful reader understanding of topic structure (Hyönä & Lorch, 2004; Murray & McGlone, 1997). However, other studies have reported that such understanding was difficult (Brown & Day, 1983) or limited even for L1 readers (Hyönä et al., 2002; Hyönä & Nurminen, 2006). In the field of L2 reading comprehension, Morishima (2013) and Ushiro, Nahatame, et al. (2016, Experiment 2) also pointed out the difficulty of distant-sentence linking during EFL reading. Their studies attributed this difficulty to the limited cognitive resources available during EFL reading, due to the high consumption of resources in lower-level reading processes such as word recognition and syntactic parsing. This possibility was also confirmed by the results of the standard condition in Experiment 4. More than 80% of the think-aloud comments were occupied with analysis within sentences, both in the higher- and lower-proficiency groups. Therefore, it can be said that the Japanese EFL readers had difficulty with topic structure processing during reading due to the high allocation of cognitive resources to lower-level reading processes.

5.1.3 Representations of Topic Structure in Text Memory

Previous research has hypothesized that readers comprehend topic structure during reading and retrieve their text memory in accordance with the structure they have constructed (Lorch & Lorch, 1985). This hypothesis has been demonstrated in L1 reading research (Hyönä & Lorch, 2004; Lorch et al., 2001; Ritchey et al., 2008; Surber & Schroeder, 2007). On the other hand, the findings of the present study showed that the Japanese EFL readers had difficulty understanding topic structure during reading, as discussed in the previous subsection. Regarding memory of topic structure, the results were inconsistent among the experiments. Specifically, the participants were able to represent the topic structure in their text memory in Experiment 1, while they failed to do so in the standard condition of

Experiments 3 and 4. There are two possibilities that might explain the conflicting results.

First, the different timings of the recall task might have affected the results. Specifically, the participants answered the recall task immediately after reading each text in Experiment 1. However, there was a delay between reading and the recall task in Experiments 3 and 4. Specifically, the participants completed the recall task after reading all the texts in Experiment 3. Moreover, the English-reading proficiency test was conducted before the recall task in Experiment 4. Hence, the participants in Experiment 1 might have remembered the links between the major topics and subtopics better simply because their text memory was still clear, while it may have been faded in Experiments 3 and 4 due to the time lag. However, this possibility seemed less valid because presenting the major topics as recall cues was still effective even in the delayed recall task in Experiment 1. Therefore, the difference between the results of the experiments was likely caused by other factors.

Alternatively, it is possible that the different recall cues influenced the representations of topic structure in the participants' text memory. In Experiment 1, reader memory of topic structure was measured by comparing the subtopic recall rates when the major topic or no information was provided as recall cues. Because topic structure processing was difficult during EFL reading, the participants were more likely to link the major topics with the subtopics during the recall task, rather than during reading. The major cues were intended to provide the participants with the context needed to link the major topics with the subtopics (Lorch & Lorch, 1985), and then to assist them in linking the cue information with the subtopics through the reconstruction of their text memory. In this way, topic structure was understood in the recall task even though it was difficult during EFL reading. In the post-reading tasks, more resources were likely to be available for higher-level cognitive processes (e.g., understanding topic structure) because the resource-demanding processes needed for literal comprehension had been finished (Hosoda, 2014; Nahatame, 2013; Ushiro, Mori, et al., 2016). Indeed, whereas previous research reported that linking distant sentences was difficult during EFL reading (Morishima, 2013; Ushiro, Nahatame, et al., 2016, Experiment 2), success in this process was observed in the post-reading task (Ushiro, Mori, et al., 2016).

On the other hand, when the introductory sentences were presented as recall cues in Experiments 3 and 4, evidence of reader memory of topic structure was not observed. The participants' recall quality in Experiment 4 demonstrated that their text memory was fragmented, and they were only able to represent the links between the subtopics and supporting details from the same paragraphs. This finding was consistent with the findings of previous studies that showed that EFL readers were able to comprehend topics summarizing paragraphs (Mori, 2015; Ushiro et al., 2008) but they were not able to understand topics beyond the paragraph level (Johns & Mayes, 1990; Ushiro et al., 2008). For example, Kimura (2014) reported that Japanese EFL readers were able to identify topics within paragraphs but failed to comprehend the overall themes of expository texts.

In contrast to Experiment 1, the participants failed to represent topic structure in text memory without retrieval cues relevant to the topic structure in Experiments 3 and 4, despite more cognitive resources being available for the post-reading task than during reading. Although this finding seems to be inconsistent with the findings of previous studies that showed that EFL readers were able to link distant sentences in a post-reading task without cues relevant to these sentences (Ushiro, Mori, et al., 2016), there are three possible reasons for this difference. First, the distance between sentences to be linked may be relevant. Whereas Ushiro, Mori, et al. (2016) targeted the linking of sentences separated by four other sentences, the focus of the present study was links between the major topics and subtopics scattered across the paragraphs of entire texts. Second, understanding topic structure requires that the reader construct more links than what is required to link isolated sentences in the past research (Ushiro, Mori, et al., 2016). Specifically, although Ushiro, Mori, et al. (2016) and Hosoda (2014) explored a single link between two sentences, studies investigating reader understanding of topic structure examined multiple links between the major topics and subtopics (i.e., two or three links in the present study; e.g., Britton, 1994; Kintsch, 1998). Finally, it is also possible that the text genre may have affected the results. Ushiro, Mori, et al. (2016) adopted narrative texts that described events related to daily experiences and thus the topics were highly familiar to the readers. In contrast, the present study used expository texts that explained concepts and topics unfamiliar to the readers (Coté et al., 1998). The participants in both Experiments 3 and 4 failed to understand topic structure because it required linking the major topics with several subtopics beyond the paragraph level in expository texts describing unfamiliar subjects, despite the cognitive resources available for completion of the post-reading task.

In summary, the present study indicates that topic structure processing was difficult for the EFL readers because their cognitive resources were mostly allocated to lower-level processes, and thus they were simply not able to dedicate cognitive resources to high-level processes, such as understanding topic structure. In addition, even though more cognitive resources were available for understanding topic structure after the lower-level processes had been completed, the EFL readers were not able to link the major topics with subtopics independently in the post-reading tasks. Because it was difficult to construct multiple links across paragraphs on unfamiliar topics, the readers needed retrieval cues relevant to the topic structure to trigger their understanding of it through the reconstruction of their text memory.

5.1.4 The Effect of Educational Interventions on Reader Understanding of Topic Structure

The above sections discussed the present study’s finding that the Japanese EFL readers could not understand topic structure automatically (in contrast to the constructionist hypothesis) during reading or in the post-reading tasks. This finding is consistent with the minimalist hypothesis (McKoon & Ratcliff, 1992), which states that readers cannot establish globally coherent comprehension of texts automatically, but only strategically. Therefore, educational interventions were needed to help the EFL readers understand topic structure. Previous research has sought to support reading comprehension by giving specific reading instructions (e.g., Horiba, 2000, 2013; Kimura, 2012, 2014, 2015a; Ushiro et al., 2017) and having the readers engage in specific tasks (e.g., Merchie & van Keer, 2016; Redford et al., 2012; Stull & Mayer, 2007; Yoshida, 2012). Study 2 explored the effects of reading instructions and task engagement, adopting an outline task that explicitly directed the EFL readers to develop a better understanding of topic structure. The effects of reading instructions and task engagement on during-reading comprehension and memory in the post-reading task will be discussed in the following paragraphs.

Regarding during-reading processes, Experiment 3 demonstrated that the outline instructions led the participants to read the texts more attentively (Ushiro et al., 2017), regardless of their L2 reading proficiency. Consistent with the framework of standards of coherence (e.g., van den Broek et al., 2015), the participants adapted their standards of coherence and during-reading processes in response to the given instructions. However, their attentive reading did not contribute to successful topic structure processing while reading. In previous studies, although L1 readers flexibly adapted their reading processes to the given reading instructions (e.g., Linderholm & van den Broek, 2002; van den Broek et al., 2001; van der Schoot et al., 2010), L2 readers were not able to do so (e.g., Horiba, 2000, 2013; Kimura, 2014, 2015b) as stably as L1 readers.

The results of Experiment 4 indicated that writing outlines while reading changed the cognitive processes of the participants; however, the task was more effective for the higher-proficiency group than the lower-proficiency group. Specifically, the higher-proficiency participants selectively reread the major topics and subtopics, and then they better understood the topic structure. They were more likely to reactivate the major topics and subtopics in their minds (e.g., Hyönä & Lorch, 2004; Hyönä et al., 2002; Hyönä & Nurminen, 2006), which was needed to link distant sentences for global coherence (e.g., van der Schoot et al., 2012). However, it should be noted that the participants wrote outlines after reading the entire texts and that the topic structure was not understood until after reading, while they were writing the text outlines. Similarly, past research has also reported that L1 readers understood topic structure better when they made graphic organizers of the major topics and subtopics after reading (Ponce & Mayer, 2014).

On the other hand, although the lower-proficiency group increased rereading and comments on text structure, this did not contribute to their understanding of topic structure. Specifically, they reread the sentences non-selectively and they still had difficulty understanding topic structure. Their outline performance indicated that their comprehension was fragmented or just clustered into paragraphs (i.e., the subtopics followed by the supporting details). This finding is consistent with Kimura (2014), who reported that Japanese EFL readers failed to understand the themes of expository texts and tended to simply understand the paragraph subtopics. Like Kimura's (2014) participants, the participants of the present study might have not been proficient at linking text information beyond the paragraph level.

A comparison of the results of Experiments 3 and 4 suggests that the understanding of topic structure by the higher-proficiency participants was aided when they engaged in the outline task after they finished reading the entire texts. Regarding the former condition, in order for the participants to understand the topic structure, just giving the outline instructions was not enough; rather, they needed to actually write the text outlines while reading. This observation is consistent with activity theory, which posits that deep text comprehension (e.g., the selection and organization of information) is achieved when readers engage in productive tasks. In the case of the present study, the participants might not have perceived much need to understand the topic structure in Experiment 3 because they did not actually write the text outlines. However, they seemed to perceive much need to understand the topic structure and adapted their cognitive processes significantly when they wrote the outlines in Experiment 4. By engaging in the productive outline task, the participants were better able to identify the major topics and subtopics in the texts and organize them in a hierarchical manner.

As for the latter condition, the higher-proficiency participants might have strategically understood topic structure when writing outlines after reading, not while reading. Because writing the text outlines in Experiment 4 required the participants to select and organize information throughout the texts, they might have attempted to comprehend necessary text information before writing the outlines. However, they had to understand the topic structure while reading in Experiment 3 in a manner different from the above strategy observed in Experiment 4, which might have made it difficult for them to understand topic structure.

Furthermore, this strategy may be related to the phenomenon of limited cognitive resources during EFL reading. Few cognitive resources were available due to the burden of lower-level processes during EFL reading (e.g., Morishima, 2013), but more resources were available after finishing these activities (Ushiro, Mori, et al., 2016). Hence, the higher-proficiency participants in Experiment 4 might have concentrated on understanding the texts during reading, and they did not write the text outlines simultaneously. If they had engaged in reading and outlining simultaneously, it might have been too resource-demanding for them. After finishing reading, they wrote the outlines well because they were able to sufficiently engage in that task.

On the other hand, neither the outline writing task nor simply receiving the outline instructions effectively supported reader understanding of topic structure among the lower-proficiency participants. This is likely due to their limited L2 reading proficiency. First, the lower-proficiency participants were likely less proficient in the lower-level processes of reading comprehension than the higher-proficiency participants. Hence, they had difficulty allocating sufficient cognitive resources to the understanding of topic structure targeted by the educational interventions (i.e., the outline instructions and outline task). This observation is consistent with the findings of previous studies that showed that the amount of available cognitive resources affected reader understanding of topic structure (Hyönä et al., 2002; Hyönä & Nurminen, 2006). Hence, the limited cognitive resources caused by less proficient, lower-level reading processes likely made the educational interventions too demanding for the lower-proficiency group. Previous research has suggested that reading instructions (e.g., Horiba, 2013) and task engagement in productive tasks (e.g., Stull & Mayer, 2007) are not effective when they are too resource-demanding for the readers. Because the outline task explicitly required the participants to both select and organize the major topics and subtopics (Lorch et al., 2013; Lorch et al. 1987), the lower-proficiency participants with limited cognitive resources had difficulty with the outline instructions and outline task.

Second, the lower-proficiency participants seemed to be less able to link text information beyond paragraphs. The outline task used in Experiment 4 demonstrated that they were less skilled at organizing information beyond the paragraph level than the higher-proficiency participants. This is consistent with the findings of Kimura (2014), who reported that L2 reading proficiency influenced Japanese EFL readers' comprehension of themes across paragraphs in expository texts.

In summary, neither the outline instructions nor the outline task facilitated topic structure processing during EFL reading. However, the higher-proficiency participants better understood topic structure when they wrote the text outlines after reading. They might have strategically concentrated on resource-demanding lower-level processes during reading, focusing on the comprehension of text information necessary for their outlines. After reading, they were able to select and organize the major topics and subtopics by actively engaging in the outline task. On the other hand, neither educational intervention was effective in promoting the understanding of topic structure among the lower-proficiency participants. The possible reasons for this outcome include limited cognitive resources for higher-level processes and poor linking skills needed to connect information and ideas beyond the paragraph level.

Experiments 3 and 4 also explored the effects of the educational interventions on the participants' memory of topic structure. The results of Experiment 3 suggest that the outline instructions enhanced the participants' overall text memory, which was consistent with their attentive mindset during reading. However, the instructions did not prompt structured memory of the topic structure, as well as during-reading understanding of it, regardless of L2 reading proficiency. Moreover, during-reading understanding and memory of topic structure did not correlate significantly. This tendency is consistent with the findings of previous L2 studies (Horiba, 2000, 2013), which suggested that the relationship between reading processes and text memory is not straightforward in L2 reading.

The results of Experiment 4 demonstrated that writing outlines enhanced the memory of topic structure among the higher-proficiency participants, but not among the lower-proficiency participants. Moreover, the outline and recall performance correlated significantly for both proficiency groups. Taken together, task engagement was and was not effective for the higher- and lower-proficiency participants in the promotion of understanding of topic structure while writing outlines. However, once the participants understood the topic structure during the outline task, their comprehension was represented in their text memory, regardless of L2 reading proficiency. In fact, in Experiment 4, among the texts in which the participants demonstrated understanding of the topic structure in their outlines, both the higher- and lower-proficiency participants later recalled the structure in 74% and 78% of the texts, respectively.

Reader understanding of topic structure resulting from completion of the outline task might have been represented in their text memory in the following two ways. First, because they better comprehended topic structure by engaging in the outline task, the topic structure itself was represented more robustly in their text memory, and they could use that structure to retrieve text memory (e.g., Lorch & Lorch, 1985). Second, cognitive processes that occurred while writing the text outlines might have induced the reconstruction of text memory. By engaging in a productive task, the participants were able to select the major topics and subtopics to be included in the outlines, as well as organize them in accordance with the hierarchy of topic structure. Therefore, these processes might have enabled them to reconstruct their text memory through the linking of the major topics and subtopics.

Furthermore, in contrast to Experiment 1, the participants in Experiment 4 were able to represent the topic structure in their text memory without relevant cues. Because the participants read the texts without engaging in a specific task in Experiment 1, they were likely to have experienced difficulty with understanding topic structure while reading. Hence, they were not able to represent topic structure in the recall task after reading without relevant cues. On the other hand, because the participants in Experiment 4 deeply comprehended the texts by writing outlines, they were able to rely on their own comprehension instead of the retrieval cues. Specifically, the topic structure was more robustly represented in their text memory, which served as a guide while recalling information (e.g., Lorch & Lorch, 1985). For example, retrieval of the major topics provided access to the subtopics that were linked to them in a top-down manner. Moreover, because they engaged in selection and organization of the major topics and subtopics while writing the text outlines, the participants were likely to engage in these cognitive processes again in the post-reading task through the reconstruction of text memory.

In summary, the outline instructions failed to support the participants' memory of topic structure as well as their during-reading comprehension of it. On the other hand, writing outlines was more effective for the higher-proficiency participants and less effective for the lower-proficiency participants. However, once the topic structure was understood through the completion of the outline task, it was also more likely to be recalled later, regardless of L2 reading proficiency. Specifically, reader understanding of topic structure was more robustly represented in their text memory, and the cognitive processes that occurred while writing the text outlines might have induced the reconstruction of text memory.

5.2 Limitations of the Present Study and Suggestions for Future Research

The present study provides new insights into L2/EFL readers' discourse comprehension and the effects of educational interventions on it. However, the present study has several limitations. In this section, the limitations of the present study will be considered in terms of the participants, materials, and methodologies.

First, as for the participants, the sample size of the present study was relatively small. The small sample size might have affected the variance in the participants' L2 reading proficiency levels, and thus it did not affect understandings of the topic structure without retrieval cues or educational interventions. In addition, although the effect of writing outlines differed according to L2 reading proficiency, the difference might increase among more participants with greater differences in L2 reading proficiency. Therefore, future research should replicate the present findings with a larger sample size and a wider range of L2 reading proficiency.

Moreover, the present study adopted L2 reading proficiency at a discourse level alone as a reader factor. Because the focus of the present study was globally coherent comprehension, discourse comprehension proficiency was likely to affect it (e.g., Kimura, 2014). In fact, the results of Experiment 4 confirmed that the higher-proficiency group outperformed the lower-proficiency group in their understanding of topic structure. However, other reader factors might have also been relevant, as stated in the discussion section. For example, the lower-proficiency group was less proficient in the lower-level cognitive processes of reading comprehension (e.g., word recognition, syntactic parsing), and thus the resources available for higher level processes, such as constructing globally coherent comprehension, were limited (e.g., Morishima, 2013; Ushiro, Nahatame, et al., 2016; Yoshida, 2003). Therefore, it is necessary to consider these reader factors as well in future research.

Second, regarding the materials, the present study adopted short and simple expository texts. Specifically, the experimental texts in the present study were shorter (i.e., less than 180 words) than those used in previous L2 research on discourse comprehension across paragraphs (e.g., Kimura, 2014; Ushiro et al., 2009; Ushiro et al., 2008). The previous studies used a single text for each condition; thus, it is possible that the specific features of the texts affected the results. Therefore, the present study assigned two short texts to each within-participant condition to avoid possible effects of text features other than the research aim, as well as to avoid placing any further cognitive burden on the participants. However, it is necessary to replicate the findings of the present study by using longer texts. Because the reactivating and linking processes are likely to be more difficult with longer texts than with shorter ones, results for studies using longer texts might be different from those of the present study.

Additionally, the texts used in the present study consisted of simple topic structures. While each text used in Experiments 1 to 3 included one major topic and three subtopics, in Experiment 4 they consisted of one major topic and two subtopics. Because the present study adopted short texts for the aforementioned reasons, these texts represented a simple topic structure based on one major topic. However, L1 reading research has examined reader understanding of more complex topic structure consisting of multiple major topics and subtopics (e.g., Hyönä et al., 2002; Hyönä & Lorch, 2004; Hyönä & Nurminen, 2006). For example, Hyönä and Lorch (2004) used expository texts with 10 subtopics organized into two major topics. Understanding such a complex topic structure is likely to be more difficult; thus, this should be further explored in future research.

Finally, the methodologies adopted in the present study did not measure on-line reading processes in natural situations, requiring the participants to engage in secondary tasks. Specifically, the priming tasks in Experiments 2A and 2B required them to make specific responses to the target words presented. The sentence-by-sentence reading in Experiments 2A to 3 required key presses, and the think-aloud methodology in Experiment 4 required the participants to verbalize their thoughts. Although all were valid methodologies widely used to assess reading processes, they might have represented task-induced cognitive processes as well as natural cognitive processes. Moreover, the methodologies used in the present study might have placed an additional cognitive load on the participants, making it more difficult to understand topic structure during reading.

Therefore, further research should explore topic structure processing in more natural reading situations without secondary tasks. One such suitable methodology is the use of eye-tracking technology. Although it has been mainly used to examine word and sentence reading processes, it is also useful in the study of discourse comprehension. Its use does not require secondary tasks, and thus the readers would not be burdened with additional cognitive loads. Furthermore, the eye-tracking method provides detailed data of ongoing reading processes (e.g., duration and frequency of fixations, rereading). This method would thus provide deeper insights into topic structure processing during L2/EFL reading.

5.3 Pedagogical Implications

The present study revealed that Japanese EFL readers have difficulty understanding topic structure independently, but some types of interventions (e.g., retrieval cues, engaging in productive tasks) were shown to support such understanding. Successful reading comprehension requires that readers not only understand the individual important topics in the text but also build global coherence among these topics. However, EFL readers often fail to comprehend entire texts, paying too much attention to translating individual words and sentences. Because of such fragmented text comprehension, text memory can suffer. Therefore, EFL readers require the scaffolding of educational interventions to understand the topic structure of expository texts. The findings of the present study have pedagogical implications, which can be summarized into the following three points: (a) the relevance of lower-level cognitive processes to the understanding of topic structure, (b) the importance of retrieval cues in post-reading tasks, and (c) the effects of educational interventions on globally coherent comprehension.

The relevance of lower-level processes and appropriate texts

The findings of the present study indicate that insufficient lower-level processes were the likely cause of the EFL readers' difficulty in understanding topic structure. Compared to L1 readers, L2/EFL readers are not proficient in the lower-level processes of English, which results in a lack of cognitive resources available for higher-level processes such as building globally coherent text comprehension (Morishima, 2013; Ushiro, Nahatame, et al., 2016; Yoshida, 2003). Specifically, in Experiment 4, the participants consumed nearly 80% of their cognitive resources in lower-level processes in the standard reading situation. Therefore, teachers need to improve EFL learners' basic reading skills at the word and sentence levels. To ensure that EFL learners have the necessary cognitive resources available to understand topic structure, it is not only necessary to teach lexical and grammatical knowledge, but also to help them improve their fluency in word and sentence decoding.

At the same time, when using the outline task in training individuals to understand topic structure, teachers should select texts that will not present difficulties in terms of vocabulary or grammar. For example, the English-reading proficiency of the higher-proficiency groups in the present thesis corresponded to the second grade of the EIKEN (level B1 on the CEFR; Dunlea, n.d.), and these participants were able to understand the topic structure in Experiment 1 and write well-organized outlines in Experiment 4. This means that the experimental texts could be outlined by EFL readers at the intermediate level. Specifically, the web tool Lexile Analyzer® (https://lexile.com/) indicates that the average lexile measure (i.e., text complexity) of the experimental texts was 733.33L to 827.78L (*SD* = 98.91; 700L–800L: *n* = 10, 800L–900L: *n* = 5, 900L–1000L: *n* = 1, 1000L–1100L: *n* = 2), which is calculated based on sentence length and log word frequency. This range is close to that of EFL textbooks for senior high schools in Japan (Negishi, 2015) and lower than that of the passages used in the EIKEN second grade test (*M* = 1000L to 1100L [*SD* = 70.71]; 900L–1000L: *n* = 1, 1000L–1100L: *n* = 2, 1100L–1200L: *n* = 1). The above analysis suggests that teachers should use texts that are much easier than EFL readers' proficiency levels when their focus is on higher-level processes such as understanding topic structure (as in the outline task). Similarly, Negishi (2015) points out that textbooks are considered highly difficult when compared with EFL learners' reading ability in Japan because schools select textbooks that EFL students should learn to read. However, texts that are too complex are likely to demotivate students and cause them to pay too much attention to word- and sentence-level processes.

Retrieval cues in post-reading tasks

The findings of the present study indicate that the Japanese EFL readers failed to independently understand topic structure while reading and during the post-reading tasks. However, the results of Experiment 1 demonstrated that presenting the major topics as retrieval cues facilitated the participants' subtopic recall. As suggested by previous studies (Hosoda, 2014; Nahatame, 2013; Ushiro, Mori, et al., 2016), EFL readers are likely to engage in higher-level processes (e.g., inference generation, building global coherence) through the reconstruction of text memory in post-reading tasks. In post-reading tasks, relatively more cognitive resources are available after the resource-demanding, lower-level processes of reading are finished. However, the recall task in the standard condition in Experiments 3 and 4 demonstrated that the EFL readers were not able to understand topic structure though the reconstruction of their text memory simply by engaging in the post-reading tasks. When engaging in the post-reading tasks, they needed retrieval cues representing the major topics.

Because major topics represent the highest level of the hierarchy of topic structure, they facilitate the linking of subtopics to major topics in a top-down manner. Furthermore, the results of Experiment 1 demonstrate that the positive effect of retrieval cues appears not only in tasks immediately following reading but also in delayed tasks. Hence, teachers can use major topics as retrieval cues to confirm learners' text comprehension right after reading. In the next class, they can present the major topics to review what the EFL students learned from the texts that were read in the previous class. In addition, although the retrieval cues were found to be effective in the promotion of reader understanding of topic structure during the recall task in Experiment 1, these cues are also expected to have a positive effect on retelling tasks, another type of reproduction task. By using retrieval cues in retelling tasks, teachers can help students to understand topic structure in a communicative way.

The effects of educational interventions on globally coherent comprehension

Because Japanese EFL readers have difficulty with the understanding of topic structure, they need educational interventions. Several previous studies have sought to improve globally coherent comprehension by giving readers specific reading instructions (Kimura, 2014, 2015a; Lorch et al., 2013; Lorch et al., 1987). The results of Experiment 3 suggest that the outline instructions were not effective in facilitating reader understanding of topic structure, although it made the participants more attentive and enhanced their general text memory. In previous L2 research, reading instructions were found to be effective in some studies (e.g., Ushiro et al., 2017) but not in other studies (e.g., Horiba, 2000, 2013). Therefore, the effect of reading instructions is not stable in the case of EFL readers, in contrast to L1 readers.

To improve their reading comprehension more directly, EFL learners should engage in the productive outline task. The outline task is likely to support reader understanding of topic structure because they require the readers to select information (i.e., the major topics and subtopics) and organize it into the hierarchy of topic structure. In fact, the results of Experiment 4 suggest that writing outlines is effective for higher-proficiency readers, who will be able to selectively reread the major topics and subtopics and then link these topics.

Moreover, the higher-proficiency readers can use their understanding of topic structure during a subsequent reading activity. In Experiment 4, when the participants understood the topic structure in the outline task, they were able to use it to achieve the written recall task later on. In the actual classroom settings, teachers can use a retelling task as a more communicative type of reproduction task. In this case, many students would require planning before the retelling in order to explain the text contents clearly and coherently. Hence, the outline task might be useful for planning the retelling task and helping students organize what they will reproduce.

On the other hand, regarding lower-proficiency EFL learners, writing outlines might not be effective. One possible reason is a mismatch of readers' proficiency and task demands (e.g., Yoshida, 2012). To reduce the mismatch, teachers need to avoid extraneous processes irrelevant to text comprehension (e.g., how to layout information; Sweller, 1988) by clearly explaining how to outline texts, showing sample outlines, and allowing the students to practice prior to engaging in productive tasks. In addition, teachers should use easy texts to decrease the cognitive load placed on their students in consideration of their proficiency levels. Another possible reason is that the lower-proficiency readers had insufficient skills for organizing text information throughout the texts. In this case, teachers need to explicitly teach them effective strategies for outlining texts. For example, the higher-proficiency group in Experiment 4 selectively reread the major topics and subtopics and paid attention to the hierarchical relationships between them, which might also be effective strategies for EFL learners with lower proficiency. Teachers probably need to train their students to outline texts regularly. In Experiment 4, the lower-proficiency group was not able to successfully complete the outline task after only one practice session, suggesting that a single session was insufficient. Similarly, several previous studies showed positive effects of educational interventions repeated over the long term (e.g., Chang et al., 2002).

As mentioned above, lower-proficiency learners have difficulty outlining texts. However, this does not mean that educational interventions are not at all effective for these learners. Experiment 4 indicated that the lower-proficiency group attempted to adapt their cognitive processes to complete the outline task even though their efforts did not lead to a successful understanding of the topic structure. If the aforementioned teacher support works well, even lower-proficiency learners can outline texts well and understand the topic structure. In fact, if EFL learners can outline texts, their text comprehension is likely to be represented in their text memory, regardless of their L2 reading proficiency in Experiment 4.

5.4 Concluding Remarks

It is at the heart of discourse comprehension to construct globally coherent comprehension of texts. However, L2 or EFL teachers often put more focus on the instruction of word or sentence-level reading comprehension. Although they sometimes teach how to understand important information in texts and paragraphs, this does not appear to be sufficient. To comprehend entire texts, it is necessary not only to understand individual pieces of important information but also to grasp the relationships among them. Despite the importance of this task, L2 and EFL readers have difficulty linking pieces of important information in expository texts. Although previous studies have also demonstrated this difficulty, they have not fully revealed which cognitive processes pose the biggest problems for EFL readers and what are the causes of such difficulties. Additionally, educational interventions to support globally coherent EFL reading comprehension have not been sufficiently developed, nor have their effects been empirically demonstrated. Therefore, the present study investigated globally coherent text comprehension and the effects of educational interventions on it, focusing on linking the major topics representing the entire texts and subtopics representing individual paragraphs. Although the present study has several limitations as mentioned above, the author believes that the findings are helpful for providing further insight into globally coherent text comprehension of EFL readers and the role of educational intervention in this cognitive process.

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

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238. doi:10.1016/0010-0277(88)90020-0

Amano, S., & Kondo, K. (1999). *NTT database series nihongo no goitokusei dai ikkan tango shinmitsudo* [NTT database series lexical features of Japanese volume 1 word familiarity]. Tokyo, Japan: Sanseido.

Amano, S., & Kondo, K. (2000a). *NTT database series nihongo no goitokusei dai nana kan hindo 1* [NTT database series lexical features of Japanese volume 7 word frequency 1]. Tokyo, Japan: Sanseido.

Amano, S., & Kondo, K. (2000b). *NTT database series nihongo no goitokusei dai nana kan hindo 2* [NTT database series lexical features of Japanese volume 7 word frequency 2]. Tokyo, Japan: Sanseido.

Bohn-Gettler, C. M., & Kendeou, P. (2014). The interplay of reader goals, working memory, and text structure during reading. *Contemporary Educational Psychology*, *39*, 206–219. doi:10.1016/j.cedpsych.2014.05.003

Britton, B. K. (1994). Understanding expository text: Building mental structures to induce insights. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641–674). London, UK: Academic Press.

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, *22*, 1–14. doi:10.1016/S0022-5371(83)80002-4

Carrell, P. L. (1984a). The effects of rhetorical organization on ESL readers. *TESOL Quarterly*, *18*, 441–469. doi:10.2307/3586714

Carrell, P. L. (1984b). Evidence of a formal schema in second language comprehension. *Language Learning*, *34*, 87–112. doi:10.1111/j.1467-1770.1984.tb01005.x

Carrell, P. L. (1992). Awareness of text structure: Effects on recall. *Language Learning*, *42*, 1–18. doi:10.1111/j.1467-1770.1992.tb00698.x

Cauchard, F., Eyrolle, H., Cellier, J-M., & Hyönä, J. (2010). Vertical perceptual span and the processing of visual signals in reading. *International Journal of Psychology*, *45*, 40–47. doi:10.1080/00207590903085513.

Chambliss, M. J. (1995). Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly*, *30*, 778–807. doi:10.2307/748198

Chambliss, M. J., & Murphy, P. K. (2002). Fourth and fifth graders representing the argument structure in written texts. *Discourse Processes*, *34*, 91–115. doi:10.1207/S15326950DP3401_4

Chang, K-E., Sung, Y-T., & Chen, I-D. (2002). The effect of concept mapping to enhance text comprehension and summarization. *The Journal of Experimental Education*, *71*, 5–23. doi:10.1080/00220970209602054

Ciullo, S., Falcomata, T., & Vaughn, S. (2014). Teaching social studies to upper elementary students with learning disabilities: Graphic organizers and explicit instruction. *Learning Disability Quarterly*, *38*, 1–12. doi:10.1177/0731948713516767

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*, 1–53. doi:10.1080/01638539809545019

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. (n.d.). *EIKEN to CEFR tono kanrensei nitsuite: Kenkyu project houkoku* [The relation between EIKEN and CEFR: Research project report]. Retrieved from https://www.eiken.or.jp/center_for_research/pdf/market/report_02.pdf

Educational Testing Service. (2015). *Mapping the TOEIC® and TOEIC Bridge™ tests on the common European framework of reference for languages*. Retrieved from https://www.ets.org/s/toeic/pdf/toeic_cef_mapping_flyer.pdf

Educational Testing Service. (2017). *TOEFL ITP® overall performance descriptors*. Retrieved from https://www.ets.org/toefl_itp/research/performance-descriptors/

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Gakken. (2005). *2005 nendo ban kako-tan book tsuki Eiken ni kyu kako mondai-shu* [The collection of previous questions with a word book for the second grade Eiken test]. Tokyo, Japan: Author.

Gernsbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, *23*, 265–304. doi:10.1080/01638539709544994

Ghaith, G. M., & Harkouss, S. A. (2003). Role of text structure awareness in the recall of expository discourse. *Foreign Language Annals*, *36*, 86–96. doi:10.1111/j.1944-9720.2003.tb01935.x

Goldman, S. R., Saul, E. U., & Coté, N. (1995). Paragraphing, reader, and task effects on discourse comprehension. *Discourse Processes*, *20*, 273–305. doi:10.1080/01638539509544942

Graesser, A. C., Li, H., & Feng, S. (2015). Constructing inferences in naturalistic reading contexts. In E. J. O'Brien, A. E. Cook & R. F. Lorch (Eds.), *Inferences during reading* (pp. 290–320). Cambridge University Press.

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

Guindon, R., & Kintsch, W. (1984). Priming macropropositions: Evidence for the primacy of macroproposition in the memory for text. *Journal of Verbal Learning and Verbal Behavior*, *23*, 508–518. doi:10.1016/S0022-5371(84)90327-X

Haberlandt, K. (1994). Methods in reading research. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1–32). London, UK: Academic Press.

Hakala, C. M., & O'Brien, E. J. (1995). Strategies for resolving coherence breaks in reading. *Discourse Processes*, *20*, 167–185. doi:10.1080/01638539509544936

Hidi, S., & Baird, W. (1988). Strategies for increasing text-based interest and students' recall of expository texts. *Reading Research Quarterly*, *23*, 465–483. doi:10.2307/747644

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. Retrieved from http://nflrc.hawaii.edu/rfl/October2013/articles/horiba.pdf

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.x

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

Education in Japan), 25, 239–254. doi:10.20581/arele.25.0_239

Hyönä, J. (1994). Processing of topic shifts by adults and children. *Reading Research Quarterly, 29*, 76–90. doi:10.2307/747739

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. *Learning and Instruction*, *14*, 131–152. doi:10.1016/j.learninstruc.2004.01.001

Hyönä, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94*, 44–55. doi:10.1037//0022-0663.94.1.44

Hyönä, J., & Nurminen, A-M. (2006) Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, *97*, 31–50. doi:10.1348/000712605X53678

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.

Japan Association of College English Teachers. (2003). *JACET list of 8000 basic words*. Tokyo, Japan: Author.

Jiang, N. (2012). *Conducting reaction time research in second language studies.* London, UK: Routledge.

Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics*, *11*, 253–271. doi:10.1093/applin/11.3.253

Johnson-Laird, P. N. (1983). *Mental model*. Cambridge, MA: Harvard University Press.

Kato, D. (2014). Effects of working memory capacity on on-line goal inference generation in EFL reading. *ARELE (Annual Review of English Language Education in Japan), 25*, 223–237. doi:10.20581/arele.25.0_223

Kim, S. A. (2001). Characteristics of EFL readers' summary writing: A study with Korean university students. *Foreign Language Annals*, *34*, 569–581. doi:10.1111/j.1944-9720.2001.tb02104.x

Kimura, Y. (2012). Effects of reading goals on narrative theme comprehension. *ARELE (Annual Review of English Language Education in Japan)*, *23*, 233–248. doi:10.20581/arele.23.0_233

Kimura, Y. (2013). The process of thematic inference generation in EFL reading: Focusing on context and readers' proficiency. *ARELE (Annual Review of English Language Education in Japan)*, *24*, 125–140. doi:10.20581/arele.24.0_125

Kimura, Y. (2014). Theme comprehension in expository texts: Effects of reading goals and L2 reading proficiency. *ARELE (Annual Review of English Language Education in Japan)*, *25*, 111–126. doi:10.20581/arele.25.0_111

Kimura, Y. (2015a). Task instructions for global coherence in EFL reading: Evidence from think-aloud protocols and a recall task. *ARELE (Annual Review of English Language Education in Japan)*, *26*, 173–188. doi:10.20581/arele.26.0_173

Kimura, Y. (2015b). *Construction of globally coherent mental representations of texts in Japanese EFL learners' reading comprehension* (Unpublished doctoral dissertation). University of Tsukuba, Japan.

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction*, *7*, 161–195. doi:10.1207/s1532690xci0703_1

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*. Cambridge University Press.

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270–1328). New York, NY: International Reading Association.

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

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, *19*, 193–220. doi:10.1191/0265532202lt227oa

Lemarié, J., Lorch, R. F., Eyrollea, H., & Virbelc, J. (2008). SARA: A text-based and reader-based theory of signaling. *Educational Psychologist*, *43*, 27–48. doi:10.1080/00461520701756321

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*, 778–784. doi:10.1037/0022-0663.94.4.778

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

Lorch, R. F. (1993). Integration of topic and subordinate information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1071–1081. doi:10.1037/0278-7393.19.5.1071

Lorch, R. F., Chen, H-T., & Lemarié, J. (2012). Communicating headings and preview sentences in text and speech. *Journal of Experimental Psychology: Applied*, *18*, 265–276. doi:10.1037/a0029547

Lorch, R. F., Lemarié, J., & Chen, H-T. (2013) Signaling topic structure via headings or preview sentences. *Psicología Educativa*, *19*, 59–66.

doi:10.1016/S1135-755X(13)70011-3

Lorch, R. F., Lemarié, J., & Grant, R. A. (2011a). Signaling hierarchical and sequential organization in expository text. *Scientific Studies of Reading*, *5*, 267–284. doi:10.1080/10888431003747535

Lorch, R. F., Lemarié, J., & Grant, R. A. (2011b). Three information functions of headings: A test of the SARA theory of signaling. *Discourse Processes*, *48*, 139–160. doi:10.1080/0163853X.2010.503526

Lorch, R. F., & Lorch, E. P. (1985). Topic structure representation and text recall. *Journal of Educational Psychology*, *77*, 137–148. doi:10.1037//0022-0663.77.2.137

Lorch, R. F., & Lorch, E. P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, *21*, 261–278. doi:10.1006/ceps.1996.0022

Lorch, R. F., Lorch, E. P., & Mogan, A. M. (1987). Task effects and individual differences in on-line processing of the topic structure of a text. *Discourse Processes*, *10*, 63–80. doi:10.1080/01638538709544659

Lorch, R. F., Lorch, E. P., Ritchey, K., McGovern, L., & Coleman, D. (2001). Effects of headings on text summarization. *Contemporary Educational Psychology*, *26*, 171–191. doi:10.1006/ceps.1999.1037

Martinez, A. C. L. (2002). Empirical examination of EFL readers' use of rhetorical information. *English for Specific Purposes*, *21*, 81–98. doi:10.1016/S0889-4906(00)00029-6

McKoon, G., & Ratcliff, R. (1980). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior*, *19*, 369–386. doi:10.1016/S0022-5371(80)90267-4

McKoon, G., & Ratcliff, R. (1984). Priming and on-line text comprehension. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 119–128).

Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Merchie, E., & Van Keer, H. (2016). From text to graphical summary: A product- and process oriented assessment to explore the development in fifth and sixth graders' dynamic construction. *Learning Individual Differences*, *49*, 348–360. doi:10.1016/j.lindif.2016.05.016

Meyer, B. J. F. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Literacy Research*, *7*, 7–47. doi:10.1080/10862967509547120

Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, *21*, 121–143. doi:10.3102/00028312021001121

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234. doi:10.1037/h0031564

Mori, Y. (2015). Do EFL readers use macrorules in their mind without the summary task? Evidence from activation of main ideas in the sentence recognition task. *KATE (Kantokoshinetsu Association of Teachers of English) Journal*, *29*, 85–98. doi:10.20806/katejournal.29.0_85

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

Murray, J. D., & McGlone, C. (1997). Topic overviews and the processing of topic structure. *Journal of Educational Psychology, 89*, 251–261. doi:10.1037/0022-0663.89.2.251

Nahatame, S. (2013). Cognitive demands and the process of predictive inference generation in Japanese EFL reading. *ARELE (Annual Review of English Language Education in Japan), 24,* 173–188. doi:10.20581/arele.24.0_173

Nahatame, S. (2014). Strategic processing and predictive inference generation in L2 reading. *Reading in a Foreign Language, 26,* 54–77. Retrieved from http://nflrc.hawaii.edu/rfl/October2014/articles/nahatame.pdf

Negishi, M. (2015). Lexile measures of textbooks used in secondary and tertiary education in Japan. *ARCLE Review, 9,* 6–16. Retrieved from http://www.arcle.jp/research/books/data/html/data/pdf/vol9_3-1.pdf

Obunsha. (2009). *2009 nendo ban Eiken jyun ichi kyu zen mondai-syu* [The collection of previous questions for the pre-first grade EIKEN test]. Tokyo, Japan: Author.

Ohlhausen, M. M., & Roller, C. M. (1988). The operation of text structure and content schemata in isolation and in interaction. *Reading Research Quarterly*, *23*, 70–88. doi:10.2307/747905

Ponce, H. R., & Mayer, R. E. (2014). Qualitatively different cognitive processing during online reading primed by different study activities. *Computers in Human Behavior*, *30*, 121–130. doi:10.1016/j.chb.2013.07.054

Poynor, D. V., & Morris, R. K. (2003). Inferred goals in narratives: Evidence from self-paced reading, recall, and eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29,* 3–9. doi:10.1037/0278-7393.29.1.3

Rawson, K. A., & Kintsch, W. (2004). Exploring encoding and retrieval effects of background information on text memory. *Discourse Processes*, *38*, 323–344.

doi:10.1207/s15326950dp3803_3

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, *22*, 260–270. doi:10.1016/j.learninstruc.2011.10.007

Riley, G. L., & Lee, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing*, *13*, 173–187. doi:10.1177/026553229601300203

Rinck, M., Gámez, E., Díaz, J. M., & de Vega, M. (2003). Processing of temporal information: Evidence from eye movements. *Memory & Cognition*, *31*, 77–86. doi:10.3758/BF03196084

Ritchey, A. K. (2011). How generalization inferences are constructed in expository text comprehension. *Contemporary Educational Psychology*, *36*, 280–288. doi:10.1016/j.cedpsych.2011.03.002

Ritchey, A. K., Schuster, J., & Allen, J. (2008). How the relationship between text and headings influences readers' memory. *Contemporary Educational Psychology*, *33*, 859–874. doi:10.1016/j.cedpsych.2007.11.001

Shimizu, H. (2015). Generation of local and global bridging inferences in L2 reading comprehension. *JACET (The Japan Association of College English Teachers) Journal*, *59*, 75–92.

Society for Testing English Proficiency. (2016). *Ichiji-/niji-shiken no gouhi wa donoyouni hantei saremasuka* [FAQ: How is it determined whether test takers pass or fail the first- and second-stage tests?]. Retrieved from http://faq.eiken.or.jp/faq/show/2020

Stull, A. T., & Mayer, R. E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, *99*, 808–820.

doi:10.1037/0022-0663.99.4.808

Surber, J. R., & Schroeder, M. (2007). Effect of prior domain knowledge and headings on processing of informative text. *Contemporary Educational Psychology*, *32*, 485–498. doi:10.1016/j.cedpsych.2006.08.002

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1207/s15516709cog1202_4

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659. doi:10.1016/S0022-5371(79)90355-4

Taylor, B. M., & Samuels, S. J. (1983). Children's use of text structure in the recall of expository material. *American Educational Research Journal*, *20*, 517–528. doi:10.3102/00028312020004517

Thorndye, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, *9*, 77–110. doi:10.1016/0010-0285(77)90005-6

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318. doi:10.1006/jmla.1994.1014

Ushiro, Y., Hijikata, Y., Shimizu, M., Nakagawa, C., Koga, T., Ohno, M., & Umehara, C. (2007). Relationships between cues types in recall tests and L2 reading proficiency. *ARELE (Annual Review of English Language Education in Japan)*, *18*, 31–40. doi:10.20581/arele.18.0_31

Ushiro, Y., Hosoda, M., Mori, Y., Tada, G., Kimura, Y., Tanaka, N., & Amagai, Y. (2017). Supporting the maintenance of global coherence with situational instruction: Evidence from eye movements during EFL reading. *JACET (The Japan Association of College English Teachers) Journal*, *61*, 89–107.

Ushiro, Y., Kai, A., Nakagawa, C., Watanabe, F., Hoshino, Y., & Shimizu, H. (2009). Effects of reading perspective on EFL learners' summary writing and importance rating. *ARELE (Annual Review of English Language Education in Japan), 20,* 11–20. doi:10.20581/arele.20.0_11

Ushiro, Y., Mori, Y., Hosoda, M., Tanaka, N., Dowse, E., Tada, G., & Nakagawa, H. (2016). How do Japanese EFL readers maintain coherence in narrative memory? *ARELE (Annual Review of English Language Education in Japan), 27,* 81–96. doi:10.20581/arele.27.0_81

Ushiro, Y., Nahatame, S., Hasegawa, Y., Kimura, Y., Hamada, A., & Tanaka, N. (2014). Narrative characters' goals and EFL readers' text comprehension: Focusing on goal explicitness. *ARELE (Annual Review of English Language Education in Japan), 25,* 1–16. doi:10.20581/arele.25.0_1

Ushiro, Y., Nahatame, S., Hasegawa, Y., Kimura, Y., Hamada, A., Tanaka, N., …Mori, Y. (2016). Maintaining the coherence of situation models in EFL reading: Evidence from eye movements. *JACET (The Japan Association of College English Teachers) Journal*, *60*, 37–55.

Ushiro, Y., Nakagawa, C., Kai, A., Watanabe, F., & Shimizu, H. (2008). Construction of a macroproposition from supporting details: Investigation from Japanese EFL reader's summary and importance rating. *JACET (The Japan Association of College English Teachers) Journal*, *47*, 111–125.

van den Broek, P., Beker, K., & Oudega, M. (2015). Inference generation in text comprehension: Automatic and strategic processes in the construction of a mental representation. In E. J. O'Brien, A. E. Cook & R. F. Lorch (Eds.), *Inferences during reading* (pp. 94–121). Cambridge University Press.

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 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:10.1111/j.1467-8624.1996.tb01900.x

van den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A 'landscape' view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp.165–187). Mahwah, NJ: Erlbaum.

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–373). Hillsdale, NJ: Erlbaum.

van der Schoot, M., Horsley, T. M., & van Lieshout, E. C. D. M. (2010). The effects of instruction on situation model construction: An eye fixation study on text comprehension in primary school children. *Educational Psychology*, *30*, 817–835. doi:10.1080/01443410.2010.510600

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*, 1655–1690. doi:10.1007/s11145-011-9337-4

van der Schoot, M., Vasbinder, A. L., Horsley, T. M., & van Lieshout, E. C. D. M. (2008). The role of two reading strategies in text comprehension: An eye fixation study in primary school children. *Journal of Research in Reading*, *31*, 203–223. doi:10.1111/j.1467-9817.2007.00354.x van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York, NY: Academic Press.

Williams, J. P., Taylor, M. B., & de Cani, J. S. (1984). Constructing macrostructure for expository text. *Journal of Educational Psychology*, *76*, 1065–1075. doi:10.1037/0022-0663.76.6.1065

Williams, J. P., Taylor, M. B., & Ganger, S. (1981). Text variations at the level of the individual sentence and the comprehension of simple expository paragraphs. *Journal of Educational Psychology*, *73*, 851–865. doi:10.1037/0022-0663.73.6.851

Winograd, P. N. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly*, *19*, 404–425. doi:10.2307/747913

Wolfe, M. B. W., Magliano, J. P., & Larsen, B. (2005). Causal and content relatedness in discourse understanding and representation. *Discourse Processes*, *39*, 165–187. doi:10.1080/0163853X.2005.9651678

Yamada, H. (2003). *2003 nenndo ban EIKEN kakomon best navi ni kyu* [The collection of previous questions for the second grade EIKEN test]. Tokyo, Japan: Dai-X shuppan.

Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing*, *28*, 1071–1097. doi:10.1007/s11274-015-1903-5

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. Retrieved from https://files.eric.ed.gov/fulltext/EJ974102.pdf

Zwaan, R. A. (1999). Five dimensions of narrative comprehension: The event-indexing model. In S. R. Goldman, A. C. Graesser & P. van den Broek (Eds.), *Narrative comprehension,* *causality, and coherence: Essays in honor of Tom Trabasso* (pp. 93−110). London, UK: Routledge.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185. doi:10.1037/0033-2909.123.2.162

Appendices

Appendix 1

English Reading-Proficiency Test

Traffic Troubles

Over the past few decades, there has been a steady increase in the number of cars on the road. As a result, large cities everywhere have begun to face a common problem: more and bigger traffic jams. In central London, for example, the average speed for cars was recently said to be less than 13 kilometers per hour —slower than that of the horse-drawn carts used 100 years ago! Traffic congestion is both a waste of time and fuel, and it also produces a large amount of air pollution.

For these reasons, cities around the world have been trying to find ways to ease such congestion. Some have tried providing free buses, while others have limited the amount of parking space available. Unfortunately, these approaches have had little effect. Experts have said for many years that the best way to prevent traffic jams is to charge drivers a fee for entering the city center. This kind of plan has always been very strongly opposed by drivers. Now, however, the Mayor of London, Ken Livingstone, has introduced just such a scheme.

Under the new system, drivers in central London between 7 a.m. and 6:30 p.m. must pay a daily fee of five pounds. This fee can be paid in advance or on the day of travel. When a driver pays, the car's number plate is entered into a computer database. Over 700 cameras have been set up around the city to film the number plates of cars and check them against the database. Drivers that do not register by the end of the day are required to pay a fine. Taxis and buses do not have to pay the fee, while cars belonging to residents of the city center are given a 90 percent discount.

Many people have criticized the system. Some say that it is unfair to people with smaller incomes, and others argue that it is too complicated to work properly. However, since it was introduced in February 2003, it has worked far better than most people expected. Indeed, if it continues to operate so successfully, we can be sure that in the future many other cities will follow London's example and introduce such fees.

(1) Why was traffic moving so slowly in London?

1. Because of the growing pollution caused by cars.
2. Because of the increase in the number of cars on the road.
3. Because drivers were doing their best to avoid traffic accidents.
4. Because drivers were trying to avoid wasting fuel in the city.

(2) The plan introduced by Ken Livingstone

1. limits the number of parking spaces in the city.
2. requires every person who enters the city to pay a fee.
3. was intended to ease overcrowding on buses.
4. was based on ideas put forward by experts for many years.

(3) Why have cameras been set up around central London?

1. To make sure drivers pay the fee for entering.
2. To make sure nobody drives after 6:30 in the evening.
3. To check that cars actually belong to residents.
4. To check that people pay when they ride on public transportation.

(4) How has the scheme turned out since its introduction?

1. It has been too complicated to work.
2. It has been so successful that other cities are already using it.
3. It has worked surprisingly well.
4. It has shown that many of the criticisms people made were night.

(5) Which of the following statements about the new plan is true?

1. The plan has resulted in faster traffic speeds in many cities.
2. Drivers must pay by the end of the day they travel.
3. There is less pollution now because traffic speeds are lower.
4. People receive a discount if they register with the database in advance.

The Blue-Blooded Crab

The horseshoe crab* is one of the oldest species of animal still alive in the world today. These sea creatures developed before man —or even the dinosaurs existed. Since they have remained more or less the same since that time, they are sometimes called living fossils. Horseshoe crabs have many remarkable features. For example, they can survive for a year without food. Also, their blood is blue in color. Today, scientists are putting these ancient creatures to some very modern uses.

Scientists are particularly interested in using the enzymes* contained within the horseshoe crab's blood. When these enzymes find harmful bacteria, they react by causing the blood cells to stick together, thus allowing the animal to protect itself. The process of getting the enzymes from the creatures is relatively simple. Live horseshoe crabs are collected and some blood is removed by injecting a needle directly into their hearts. While this may seem cruel, most of the horseshoe crabs are safely returned to the ocean. The enzymes are then removed from the blood and freeze-dried into a powder. This powder is used in a procedure known as the Limulus test.

The Limulus test has been used by scientists since the 1970s. Medical professionals regularly use it to check equipment, such as surgical instruments. Water is poured over an instrument and then mixed with the powder made from the blood of the horseshoe crab. Just as they do in the creature's body the enzymes in the powder react if they come into contact with bacteria, which would show that the instrument is not safe to use. The Limulus test is also now used by NASA, America's space agency, to make sure that spacecraft do not carry any bacteria into space.

In the future, NASA may also use this test when exploring other planets. If the enzymes were to react, then this would be a sign of life on the planet. In this way, one of the oldest animals on our planet might end up helping us to make new discoveries in outer space.

* horseshoe crab: カブトガニ *enzyme: 酵素

(6) Horseshoe crabs

1. are too old a species to be used in modern science.
2. must eat once a year for their blood to stay blue.
3. first appeared at the same time as the dinosaurs.
4. have remained basically unchanged over time.

(7) What do the enzymes in the horseshoe crab's blood do?

1. They make it easy to remove the blood.
2. They stop bacteria from harming the animal.
3. They protect the blood by changing its color.
4. They help the animal to return safely to the ocean.

(8) One current use of the Limulus test is

1. to check how many enzymes are in the horseshoe crab.
2. to find other uses for medical equipment.
3. to discover whether spacecraft have any bacteria on them.
4. to make sure that water is safe for the horseshoe crab to drink.

(9) In the future, NASA may use the Limulus test

1. to take bacteria safely to other planets.
2. to find out whether there is life on other planets.
3. to enable them to protect the oldest animal on our planet.
4. to make it easier to discover other planets.

(10) Which of the following statements about horseshoe crabs is true.

1. They are called living fossils because they can survive without food.
2. Their fossils are collected, freeze-dried, and made into powder.
3. The Limulus test can be used to hold their blood cells together.
4. These ancient creatures can help us make future discoveries

Digital Witness

After a traffic accident, it is often difficult to determine exactly what happened or who was to blame. This is because people who witness an accident often describe it in different ways. In order to solve this problem, a special device has been developed that records information while a vehicle is moving. By using this information, it becomes possible to learn exactly how and why an accident occurred.

This device is actually based on the famous "black boxes" that have been used in airplanes since the 1950s. It consists of a small computer that continuously monitors the speed of the vehicle and other data. When the device senses an unusual movement, such as a sudden stop, it automatically stores the driving data. Thus, if an accident occurs, the data can later be examined to learn more about what happened.

Some devices also include a digital video camera that records the view from the driver's seat. This provides a picture of what the driver sees during an accident. The video evidence can show which person is responsible for the accident, allowing drivers to avoid costly court cases. * But such video recordings are not only useful for determining the cause of an accident. They can also help people learn to drive more carefully by showing them the mistakes they have made. For these reasons, some insurance companies even give discounts for vehicles that use this type of device.

Because of their many advantages, the devices are quickly being adopted in several parts of the world. Some bus and truck companies in the United States have started to use them in their vehicles. In London, police cars are being fitted with the devices to encourage police officers to drive more safely. As this new technology becomes more common, experts predict that it will help prevent the accidents that made such devices necessary in the first place.

*court case: 訴訟 (裁判上の争い)

(11) Why is it difficult to determine the cause of a traffic accident?

1. Drivers often refuse to describe the details of the accident.
2. People do not always agree about what happened.
3. Most vehicles move too fast to record information.
4. The new devices in cars do not provide enough information.

(12) When an accident occurs, the device

1. causes the car to stop automatically.
2. saves data about what happened.
3. changes the speed and movement of the car.
4. uses data from airplanes to monitor driving.

(13) What is one advantage of devices that include digital video cameras?

1. They can be bought cheaply from insurance companies.
2. They can show the driver which other vehicles have the devices.
3. They can make it unnecessary for drivers to have car insurance.
4. They can be used to determine who is responsible for an accident.

(14) As more vehicles are fitted with the device, it is expected that

1. bus companies will be able to hire more drivers.
2. officers will no longer have to think about safety.
3. the number of accidents will decrease.
4. the police will take more drivers to court.

(15) Which of the following statements is true?

1. There were few accidents in the U.S. before the device's introduction.
2. The idea for the device came from police cars in London.
3. The device shows drivers the shortest way to their destination.
4. The device is similar to those used in airplanes since the 1950s.

Spider Silk

Spider silk is the name given to the fine thread that spiders produce to make the webs in which they live. For many years, scientists have been impressed by spider silk’s amazing qualities. Although it is 10 times thinner than human hair, it is five times stronger than steel of the same thickness. If such a strong, light thread could be produced in large quantities, it would provide many benefits for humans. Unfortunately, spiders usually attack each other if they are kept close together. For this reason, raising large numbers of spiders together to produce thread has proved very difficult.

As a result, scientists have turned their attention to producing spider silk by other methods. The proteins* that make up spider silk were first is covered more than 10 years ago. Spider silk consists of two different proteins that the spider produces in its body. However, finding a method to produce these proteins in a laboratory has not been easy. Recently, though, scientists have made remarkable progress toward solving this problem by using genetic engineering.* For example, a group of scientists in Quebec, Canada, removed from spiders the genes that are responsible for producing thread. By placing these genes in the cells of female goats, the scientists succeeded in raising goats that can produce milk containing the proteins. They then used the proteins in the goats' milk to produce a thread very similar to that which spiders produce naturally.

Based on these results and those of other experiments, scientists now predict that man-made spider silk will soon become available for commercial use. Besides being light and strong, the new thread has other advantages. Since it is based on a natural product, it can be made without using dangerous chemicals. Moreover, the thread itself does no harm to the environment. Unlike most man-made materials, after spider silk has been thrown away, it disappears naturally over time. In the future, it may be used to make such things as fishing nets, clothing, and medical materials. There are still many problems to be overcome before spider silk can be produced cheaply, but this amazing thread promises to be useful in a wide variety of fields.

*protein: タンパク質　*genetic engineering: 遺伝子工学

(16) Why is it difficult to use spiders to produce large quantities of thread?

1. It takes spiders too long to produce the thread.
2. The thread produced by spiders is too thin.
3. Scientists are often attacked by spiders.
4. Spiders cannot easily be raised together.

(17) Recently, scientists in Quebec

1. used the genes of spiders to create thread.
2. discovered the proteins that make up spider silk.
3. produced thread by giving goat milk to spiders.
4. placed special chemicals into the milk of goats.

(18) What is one advantage of man-made spider silk?

1. It can be used to get rid of dangerous chemicals.
2. It can be created without using harmful chemicals.
3. It does not disappear naturally like other materials.
4. It does not require any natural products to make.

(19) Today, man-made spider silk

1. is still expensive to produce.
2. is found in a wide variety of products.
3. is easier to produce than other materials.
4. is not yet strong enough for commercial use.

(20) Which of the following statements is true?

1. Scientists produced thread by putting proteins into spiders' bodies.
2. Scientists were not aware of the benefits of spider silk until recently.
3. Spider-silk proteins taken from goat milk were used to make thread.
4. Spider silk has already replaced natural materials in many fields.

Moon Tales

What do you see when you look up on the night of a full moon? Westerners might gaze at the earth's only natural satellite and recall the old myth about the moon being made of green cheese. Japanese have traditionally seen a rabbit pounding rice cakes on the moon's cratered face. But for one 19th-century U.S. newspaper, the heavenly body resembled a big dollar sign.

On the morning of August 25, 1835, New Yorkers were greeted by the following news on the front page of the *New York Sun*: Great Astronomical Discoveries Lately Made by Sir John Herschel. In a series of six fictional articles, the newspaper claimed that the famous British astronomer had discovered something amazing about the moon while observing the skies at the Cape of Good Hope. Herschel had reportedly used a "telescope of vast dimensions and an entirely new principle" to sight not only trees, oceans, and beaches on the moon, but herds of goats and buffalos as well. In exaggerated language, the articles described even more incredible lunar inhabitants: blue unicorns, two-legged beavers that lived in huts, and furry, winged humanlike creatures.

The *Sun*'s publisher, Benjamin Day, claimed daily circulation jumped from 15,000 copies at the beginning of the series to 19,360 the highest in the world at that time—when it reported what would later be known as the Great Moon Hoax. Remarkably, rival papers claimed they too had obtained copies of the original articles and they rushed to reprint the series. Accounts of the public's reaction to the articles differed. According to a researcher writing in 1854, students and professors at Yale College at the time read the *Sun* excitedly and had no doubts about the truth of the revelations. Another New York paper, however, was amazed that "any man of common sense could read the stories without at once realizing the deception."

Herschel was not aware of the articles until much later and laughed them off. Privately, some Sun staff admitted to the possibility they had fooled readers, but the paper refused to publicly acknowledge any trickery. The author of the articles is generally believed to have been Cambridge-educated reporter Richard Adams Locke; he may have wanted to ridicule earlier claims by intellectuals, such as Thomas Dick, that the moon was populated. Whatever his motivation, Locke's tales are thought to have permanently raised the circulation of the Sun, which continued publication until 1950.

Looking back at the Great Moon Hoax from the 21st century, it may seem harmless enough and even irrelevant to us now. Nonetheless, in this age of 24-hour cable news channels and the Internet, all too often people are willing to believe what they see or hear without question, a dangerous trend in these dangerous times. Perhaps another media deception like the Great Moon Hoax would remind the public to be a little more cautious.

(21) What claim was made in the six articles published by the New York Sun in 1835?

1. Sir John Herschel's telescope was not powerful enough to see all the strange moon creatures he mentioned.
2. Sir John Herschel's observations were not the first to prove that various animals on the moon existed.
3. Sir John Herschel's descriptions of strange lunar creatures and landscapes would be believed by no one.
4. Sir John Herschel's work showed that a variety of life forms had been discovered on the moon.

(22) When the series first appeared in the Sun,

1. it was clear that most people were highly doubtful of the articles' claims.
2. other newspapers decided to publish the stories even though they were not true.
3. Herschel contacted the Sun and said it was wrong to use his name to trick the public.
4. journalists were afraid that readers would no longer believe the stories newspapers printed.

(23) What was one effect the Great Moon Hoax had on the Sun newspaper?

1 It damaged the reputation of the newspaper and later led to a lower readership.

2 It served to boost the popularity of the newspaper in both the short and the long term.

3 It permanently damaged Herschel's reputation as a science writer for the newspaper.

4 It encouraged readers to research more about the moon story and other strange articles.

(24) In conclusion, the author implies that

1. more precise media reporting would probably lead to increased readership of newspapers.
2. pressure to increase profits is the main reason for the faulty reporting that often occurs on Internet news sites.
3. people frequently fail to think carefully about the accuracy of stories they read in the media.
4. journalists need to be more careful to make sure the information they get is authentic.

Teaching Nomads to Read

In some countries, spreading literacy among nomadic ethnic groups has been a major challenge. Not only are the logistics difficult, the process can also involve actually changing the way people think. That has been the case with the efforts of one non-governmental organization (NGO) to spread literacy among the Van Gujjars, a nomadic ethnic group in India that lives in the forests of the Himalayan foothills. As they migrate from one region to another, they and their children remain outside the state-run educational system for seasons at a time. This has become a serious problem, as their inability to read documents or contracts makes them vulnerable to exploitation by loggers or businesspeople wishing to purchase land for development.

Hoping to change that, one NGO, the Rural Litigation and Entitlement Kendra (RLEK), has, in cooperation with the Indian government, been attempting to reach these nomads and teach them to read. This special literacy program, established in 1992, began with a group of Van Gujjars aged 15 to 35. Over the years, RLEK has dispatched about 350 volunteer teachers who have lived and traveled with the tribe, ensuring that they receive a year-round education. Maintaining close and friendly contact with the tribe is essential, but has not been easy: the Van Gujjars have long been suspicious of outsiders. As one tribe member explains, "Our children should remain behind the veil of the forest; otherwise they will acquire the evils of urban society."

The program depends on highly dedicated volunteer teachers who use innovative materials. However, it's no stroll in the park for them. They must migrate everywhere with the nomads, trekking huge distances and enduring harsh living conditions without modern comforts or conveniences. RLEK's learning materials are specially designed and contain contexts familiar and relevant to the Van Gujjars. Mobile libraries, which travel among the tribe's scattered communities, constitute another important part of the program.

Another problem has been India's forestry department, which sees the nomads as a threat to the ecology in the regions where they migrate. The problem stems from the Van Gujjars' livelihood, the tending of buffalo herds as they roam over wide areas. The tribe earns its money by selling the animals' milk. This job dictates the Van Gujjars' nomadic lifestyle. Many foresters claim that the Van Gujjars and their roaming herds have seriously damaged forests. Officials have even attempted to have RLEK teachers arrested, claiming that educating the nomads is illegal. Yet another obstacle has been local governments. One official has even tried to force the Van Gujjars to settle in designated areas. The nomads, however, have strongly resisted. "If you are settled, you are like a stone," a Van Guj[jar elder said.

No one is certain what the future holds for the nomads, but RLEK hopes its efforts will open doors for them and provide them with a broader perspective to make the best decisions for themselves and future generations.

(25) One reason the Van Gujjars have received less education than other groups is that
1. the Indian government has been unwilling to include them in formal schooling.
2. their lifestyle prevents their children from participating fully in ordinary schools.
3. the state-run educational system cannot afford special programs for them.
4. developers and loggers have threatened government teachers who try and enter Van Gujjar territory.

(26) In trying to apply its literacy program, RLEK has
1. had problems recruiting enough teachers willing to teach the Van Gujjars.
2. realized it will not receive the support it needs from the Indian government.
3. had to deal with the nomads' efforts to protect their children from mainstream society.
4. found that the Van Gujjars frequently refuse to let teachers travel with them.

(27) What special effort has RLEK made to help the nomads?
1. It has fought against the Indian government's plans for the future of the tribe.
2. It has convinced them that their children should be allowed to stay in state-run schools.
3. It has tried to persuade them to give up their traditions and settle in one place.
4. It has created educational materials that more appropriately fit their lifestyle.

(28) Which of the following is true of the Van Gujjars?
1. They feel that the action taken against RLEK by India's forestry department has been unfair.
2. They feel that it is of great importance for them to maintain their nomadic lifestyle.
3. They are dissatisfied with RLEK's efforts to educate them and have decided to stop cooperating.
4. They have found that the best way of improving their lives is to cooperate with local government officials.

Appendix 2
Experimental Texts and Yes–No Comprehension Questions Used in Experiments 1, 2A, and 2B

The major topics are boldfaced, and the subtopics are underlined. The titles were not presented in the experiments. The yes–no comprehension questions were used in Experiments 2A and 2B.

Distance

There are elements for measuring distance. <u>Economic distance is changed by the cost of movement from one place to another</u>. Money and energy are related to any movement. Sending something by water is usually less expensive than sending over land. This holds true even when land routes are shorter.

<u>Distance can be measured on the basis of time</u>. Some maps use travel time instead of mile signals. This is because the measuring unit influences the usual relations among locations. It may take the same time to go from a single point to a location 10 miles north as going to a location 30 miles south.

<u>Distance measuring varies with individual feelings</u>. What may seem like a long trip to some individuals may seem short to other people. Even the same route going and coming can seem different to a single traveler. It depends on whether road conditions are good, and whether the trip is near the end.

Q.1 海上輸送は陸上輸送より安い　Q.2 地図に移動時間が載っている
Q.3 誰もが旅行を長く感じる

Argentina

Argentina is a country with unique points. <u>A huge land called the Pampas spreads across Argentina.</u> The flat land is called the Pampas. Cowboys take care of a lot of cows on the farms. The farms are as large as those in Texas. In the place where rivers flow into the Atlantic, the capital city spreads with elegant buildings and broad roads, almost like in Paris.

As for economy, Argentina has few natural resources. The country has almost no iron or other minerals. Thus it cannot manufacture products, such as cars, machinery, and clothing. Thus, these items must be imported from other countries.

In Argentina, the society has experienced development been upgraded in a modern way. The streets in the big cities are crowded with busy people. In recent years, the newspapers and magazines have been made free to publish everything. This is because the people of Argentina were finally able to establish a liberal government.

Q.1 首都はパリのようだ　Q.2 アルゼンチンには大量の鉄がある
Q.3 アルゼンチンの人々は自由な政府を打ち立てた

Environment

The United States has rules on environmental damage. There are rules for the air. As emission gas released into the air causes serious health problems to many people, some factories must close when air condition becomes bad. Some laws keep new factories from being built in places where air condition is bad, while laws have been made to stop burning of waste, leaves, and other things.

Laws have been protecting clean water. Since people need safe water to use, some laws have stopped homes, factories, and cities from pouring dirty water into rivers and lakes. Other laws have made farmers keep insect killer from being poured into rivers and lakes when it rains.

Some laws are targeted at thrown away wastes. Because many people throw away a lot of things these days, most cities and towns have laws against dumping wastes in any place. Laws also require towns and cities to have pick-up services.

Q.1 大気の状態が悪いと工場を閉鎖しなくてはならない
Q.2 家庭は汚水を川に流すことができる
Q.3 多くの市ではごみのポイ捨てを禁止している

Support

Support for developing countries causes problems. Some kinds of support have negative effects rather than good effects. Money or machinery is given to a developing country on certain conditions. These mean that the receiving country has to buy the products of the giving country. The developing country may have to buy unnecessary or expensive products.

Support causes dependence of makes developing countries dependent on developed countries. For example, a developing country may be given expensive tractors. Farmers may produce more crops than before. However, when the tractors are broken, they will require expert mechanics or expensive spare parts. Either way, the poor country needs to pay money to the rich country.

Support for cities causes population concentration. This makes city life look more attractive. It offers highly paid jobs. Such jobs are not available in rural areas. Thus, people leave the countryside for cities. As a result, the countryside becomes empty. In contrast, the cities become crowded.

Q.1 途上国は先進国から安い製品を買うことができる
Q.2 途上国は専門の技術者を必要とする
Q.3 給料の高い仕事は途上国の田舎にもある

Energy

Electronic energy production from oil has major problems. The danger of producing electric energy is increasing. There is not as much oil left as before. It is necessary to drill even deeper to recover oil. It means greater danger to the workers. The development of nuclear* power reflected the limited amount of oil.

The decreasing availability has made oil price more expensive. It costs more money to produce oil. This situation puts the oil seller in a powerful position. Increasing costs of oil are the strongest motivation to seek alternative energy.

Our energy habits have damaged the environment. Burning oil sends out various particles to our atmosphere. Such particles include ozone and other gases. In the US, nearly half of the population lives in areas which have dirty air. Moreover, accidents of oil tankers did a lot of damage to ocean ecologies. One example is the tanker accident in Alaska.

*nuclear: 原子力の

Q.1 石油採掘の労働者に対する危険は減少している

Q.2 石油価格の高騰は代替エネルギー探究の動機を減少させている

Q.3 アラスカでタンカー事故が起こった

Three Mile Island

The accident at Three Mile Island has had wide effects. The accident has increased public consciousness sensitivity about the risks. In the U.S.A, a special committee has established strict standards to protect public health and safety. In the world, resistance to nuclear* power programs has increased. Many people have demonstrated against the nuclear energy use.

The accident influenced economy had economic impacts. It cost a lot of money to repair the power plant. The accident also affected both local and international markets of nuclear power plants. The costs of nuclear power plants have increased. This is because it has become necessary to deal with additional safety measures.

The accident has shown that there is no perfect technology and no security against human error. This understanding has resulted in a demand for the system to be changed. Nuclear plant operators should have more training. They have to handle even the most impossible troubles.

*nuclear: 原子力の

Q.1 原子力に対する反対は減少している　Q.2 原子炉の価格は高騰している

Q.3 原子力発電所の作業員は起こりえない問題にも対処する必要がある

Appendix 3
Experimental Texts and Yes–No Comprehension Questions Used in Experiments 3 and 4

The major topics are boldfaced, and the subtopics are underlined. The titles were not presented in the experiments. The Chimpanzee text, Energy text, Inventor text, and Peru text were used in Experiment 4 after the italicized sentences were removed. Because of this change, the yes–no comprehension questions of the Energy and Inventor texts were revised in Experiment 4.

Chimpanzee

We often see TV programs that show wildlife. **Compared with other animals, chimpanzees have an interesting way of life.** *Chimpanzees find the plains as an ideal territory. They spend their day walking around the wide open spaces where the tall grasses hide them while they play. The chimpanzees build nests in the trees at night because the trees provide safety from most of their enemies.*

Chimpanzees live in a complex society. Usually, five or six males lead females and their babies. The adult males provide protection for the group while the females are responsible for the babies. The adult males occupy a dominant position within the group. Further, fights often establish hierarchy within males.

Chimpanzees communicate mainly with hand signals. Since chimpanzees live in open spaces, this system is the safest way to communicate. One reason is that hand signals are silent. Even if an enemy is found, chimpanzees can inform each other of the danger without making noise and take cover to avoid being discovered.

Q. チンパンジーのオスは子どものころから群れを守っている

Energy

These days, more and more energy is being produced by natural forces such as sunshine, wind, and waves. **It is said that electronic energy production from oil has big problems.** *The danger of producing electric energy is increasing. There is not as much oil left as before. It is needed to drill even deeper to recover oil. It means greater danger to the workers. The development of nuclear power reflected the limited amount of oil.*

The decreasing availability has made oil more expensive. It costs more money to produce oil. This situation puts the oil seller in a powerful position. Increasing costs of oil are the strongest motivation to seek alternative energy.

Our energy habits have damaged the environment. Burning oil sends out various particles to our atmosphere. Such particles include ozone and other gases. In the USA, nearly half of the population lives in areas which have dirty air. Moreover, accidents of oil tankers did a lot of damage to ocean ecologies. One example is the tanker accident in Alaska.

Q. 石油価格の高騰は代替エネルギー探求の動機としては弱い (Experiment 3)
Q. 石油価格の高騰は代替エネルギーを探究する強い動機である (Experiment 4)

Inventor

If Thomas Edison had not been born, his inventions, such as the telephone, would not exist. **History shows all great inventors have common characteristics.** *Successful inventors need a strong interest in their invention. They are usually quite interested in what they are doing. They may not notice that many hours have passed. They want to know how they can overcome a difficulty. They look forward to seeing their inventions surprise many people.*

Inventors work very hard to improve their inventions. They work like busy bees. Inventors can work all day and night without meals. They know they have to make machines that work without any trouble. If their machines do not work well, nobody will use their inventions. Moreover, if their machines cause serious problems, they may injure many people.

Inventors need the ability to learn on their own. Inventors do not have to be experts who know everything. However, they discover their findings with repeated experiments. For example, Thomas Edison always experimented in his laboratory, creating a total of 1,093 inventions.

Q. 発明家は困難の克服方法を知りたいと思っている (Experiment 3)
Q. 発明家は昼夜を通して働く (Experiment 4)

Peru

Peru is a country next to Brazil, and it is famous for soccer. **The development of Peru has been influenced by its characteristics as a country.** *As for the geography, the Andes Mountains cover most of Peru. The top of the mountains are covered with snow. Even the valleys are so high that the air is thin and cold. In the areas where the mountains reach down to the Pacific Ocean, it never rains and the air is so dry.*

Industry of Peru is famous for traditional craft. The Indians who live in the high mountain valleys take care of sheep. The women make beautiful sweaters and blankets from wool, just as their ancestors did. These products have eventually become popular in Europe and the United States.

The society in Peru has kept its traditional style. When the Spanish destroyed the Inca empire, the Indians became the poor and exploited in their own land. Although a small group of White farm owners from Spain have ruled the country ever since, the great Indians remain outside the Spanish culture.

Q. スペインの白人農場主に支配され、インディアンの伝統は途絶えてしまった。

Film

In the early 1930s, less people went to the theater because the economy was bad worldwide. **By the mid-1930s, new types of film were born, and people returned to the theaters.** Documentary films reported events in the world. The bad economy changed mind of people. A lot of people wanted to understand more about the world beyond their immediate neighborhoods. Documentaries were an interesting and educational way for people to get information.

Gangs became a popular movie topic during the 1930s. These types of movies reflected the shocking newspaper headlines of the day. Actions were played out on the city streets. The streets were very familiar to movie audiences.

Audiences liked to see society saved from fear in horror films. Horror movies were first shown in the 1930s. Two popular classic horror films are “Frankenstein”" and “Dracula.” In 1930s, the ticket system of movies changed into the bonus feature. Audiences got two movies for the price of one. This discount encouraged them to watch movies.

Q. 多くの人々は自分のすぐ近くのことを超え、世界のことを知りたいと思っていた

Heart disease

Modern people have health problems due to irregular sleep cycle and lack of exercise. **Since the mid-1960s, medical progresses have been made in dealing with heart diseases.** Researchers have developed new drugs for heart diseases. For example, many doctors use chemicals called “beta blockers.” Beta blockers can decrease the blood pressure of people with high blood pressure.

The medical techniques used in operations have improved greatly. Today, doctors are able to replace not only damaged parts but the entire heart. If a heart no longer works, it can be replaced with a healthy heart from someone who has died. In other cases, mechanical devices, such as pacemakers, can be implanted in bodies to keep hearts working.

Doctors can prevent heart diseases better because they have learned more about the causes. Scientists have shown that diet can be an important means of controlling heart disease. For example, the fat cholesterol is known to block blood flow. Similarly, salt is known to increase blood pressure. Therefore, doctors emphasize the importance of a diet low in cholesterol and salt.

Q. 医者は病にかかった部分だけでなく、心臓全体を移植することができる

Nationalism

There is much in history that divides the Middle Ages from the Modern Age. **One difference is nationalism, which was created by important ideas.** Geography has divided separated areas into different nations. This is clear in ruling and ruled countries far away from each other. For example, America was ruled by Great Britain across the ocean. The Americans came into conflict with Great Britain in terms of economy. Finally, War of American Independence occurred.

The rulers and the ruled people often believed in different religion. One example is the Christians under Turkish rule in East Europe. Another example is Islam under British rule in North Africa. These peoples began to resist foreign rule. Religious background caused the feeling of being different.

Language has divided people into different ethnic groups. There used to be people called "Balkans." In the single ethnic group, some language groups were included. Thus, Balkans broke up into Greeks, Serbs, Romanians, and Bulgarians. In the case of Arab nationalism, people speaking Arabic were regarded as one nation.

Q. 1 つの民族に複数の言語グループが含まれることがある

Support

More and more, mass media are reporting about the support given from rich countries to the poor ones. **Support for developing countries often causes problems.** Some kinds of support have negative effects rather than good effects. Money or machinery is given to a developing country on certain conditions. These mean that the receiving country has to buy the products of the giving country. The developing country may have to buy unnecessary or expensive products.

Support makes developing countries dependent on developed countries. For example, a developing country may be given expensive tractors. Farmers may produce more crops than before. However, when the tractors are broken, they will require expert mechanics or expensive spare parts. Either way, the poor country needs to pay money to the rich country.

Support for cities causes population concentration. This makes city life look more attractive. It offers highly paid jobs. Such jobs are not available in rural areas. Thus, people leave the countryside for cities. As a result, the rural areas become empty. In contrast, the cities get crowded.

Q. 途上国は先進国から安い製品を買うことができる