

**THE EFFECTS OF TYPES OF QUESTION ON EFL LEARNERS'
READING COMPREHENSION SCORES**

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Education

By
Kazuhiro Ehara
August, 2008

©

Copyright

2008

by

Kazuhiro Ehara

ABSTRACT

Little empirical research has been conducted on what effect task-based reading instruction with reading questions will have on reading comprehension, particularly in the domain of second language reading comprehension.

The purpose of this research is to investigate which type of questions, textually explicit (TE) or inferential (IF) questions, will best facilitate text comprehension, and which type will have the most beneficial effect on Japanese EFL learners at three proficiency levels (low, intermediate, and high). In the study, two groups of Japanese senior high school students ($N = 69$) were classified into three different proficiency groups.

One group received instruction emphasizing TE questions while the other received instruction emphasizing IF questions. TE questions are text-bound questions whose answers are locally and explicitly stated in the text. In contrast, IF questions are more knowledge-bound questions whose answers largely depend on readers' cognitive resources, such as relevant linguistic knowledge, background knowledge, world knowledge or context. The different treatments lasted five months. The results were statistically analyzed.

The study revealed a significant task effect for reading questions on Japanese EFL learners' reading. Although one type of instruction did not have a significantly better effect than the other, the large between-groups gain gap seems to imply that instruction emphasizing IF questions might facilitate text comprehension more. The study also found that the participants who received instruction emphasizing IF

questions benefited from their instruction regardless of proficiency level. With regard to instruction emphasizing TE questions, the higher proficiency participants benefited significantly more from their instruction than the lower proficiency students.

The study suggests that reading teachers should use a task-based teaching method with reading questions. If the use of reading questions is already a part of reading teachers' methodology, they should include not only commonly used textually explicit reading questions but also inferential ones. The study suggests that implementing these changes might help break the cycle of translation-bound reading instruction with its overemphasis on lower-level processing, and might lead students to read texts in a more meaningful, interactive way.

ACKNOWLEDGMENTS

I would like to express deep appreciation for the support and assistance I received from many people in planning, conducting and reporting this research project. Without the cooperation and great patience of the following people, this study could not possibly have been completed.

First of all, I would like to thank the members of my defense committee, Drs. Edward Schaefer, J. D. Brown, Kenneth Schaefer, Mark Sawyer, and Marshall Childs, who all patiently read the manuscript and provided detailed comments, suggestions and advice.

I would like to express my sincere gratitude to Dr. J. D. Brown who taught me testing and statistics, and provided guidance on my research project and encouragement throughout the entire program of graduate studies. I am most indebted to him for his sharp insight into and valuable suggestions for research design, data collection and analysis, results, and discussion.

I wish to express my heartfelt gratitude to Dr. David Beglar for his illuminating suggestions and constant support all through the Dissertation Writing course. The writing of this research would not have been possible without his strenuous efforts to read rough versions of this manuscript and he repeatedly provided encouraging words on my work until the Oral Examination day.

I would also like to express special thanks to Dr. Edward Schaefer for patiently reading a revised version of this manuscript and providing me with useful advice and informative comments.

My heartfelt gratitude goes to Dr. Kenneth Schafer for being always supportive and helpful, and giving me a driving force to keep working. I will never forget his positive and spirit-raising words whenever I was in trouble.

I would like to thank Edward Torrico for helping to decide on model answers to the short-answer tests, proofreading the rough version of this manuscript, providing me with many important and insightful suggestions, and being supportive all the time.

I am deeply grateful to Prof. Robert Juppée for his assistance in determining model answers to the short-answer tests and marking the answer sheets to check interrater reliability. I am also thankful to him for proofreading this manuscript, providing me with many useful suggestions, and offering me constant encouragement.

I would like to acknowledge the cooperation of the senior high school students who participated in the pilot and the main studies.

I also wish to thank my colleagues for their assistance in conducting the pilot studies, and their understanding and support as I worked on this degree. I am greatly indebted to my parents, Takashi and Michiko, for giving me encouraging words. Finally, I would like to express special thanks to my wife, Rieko, who gave me the opportunity to pursue further study at Temple University and helped me to

construct tables and figures. Her constant support for the past several years enabled me to keep working and complete this research.

TABLE OF CONTENTS

	PAGE
ABSTRACT	iv
ACKNOWLEDGMENTS	vi
LIST OF TABLES	xii
LIST OF FIGURES	xiv
CHAPTER	
1. INTRODUCTION	1
The Background of the Issue.....	1
Statement of the Problem.....	3
Teachers' Approaches to Reading Instruction.....	3
Japanese Students' Attitude towards Reading.....	5
Purposes of the Study.....	7
The Audience for the Study	7
Delimitations	8
The Organization of the Study	8
2. REVIEW OF THE LITERATURE	9
What Processes Does Reading Involve?	9
Functions of Reading Questions	12
Taxonomies of Reading Questions	14
Lower-order and Higher-order Questions	16
Types of Reading Questions Based on Reading Skills.....	16
Types of Reading Questions Based on Levels of Comprehension and Reading Abilities	20

Types of Reading Questions Based on Question and Answer Relationships	21
Convergence and Divergence Reading Questions.....	23
Empirical Studies on Japanese EFL Readers and Reading Test Questions	24
Taxonomies of Inferences Involved in Reading	26
A Taxonomy of Inferences Based on Information Source	27
A Taxonomy of Inferences Based on Generation Time.....	32
Summary of Literature Review	38
Research Questions	40
3. METHODS	42
Participants	42
Instrumentation	45
Tailored Cloze Tests	45
Short-answer Tests	48
Research Design.....	51
Question Types	52
Procedures	54
4. RESULTS	57
Data Analysis	57
Test Administration Time and Treatment Effects on Reading Questions.....	57
Test Administration Time and Treatment Effects on Textually Explicit and Inferential Reading Questions.....	58
Test Administration Time and Treatment Effects on Textually Explicit Reading Questions	62

Test Administration Time and Treatment Effects on Inferential Reading Questions	67
Proficiency Effects on Reading Questions.....	71
Proficiency Effects on Textually Explicit and Inferential Reading Questions.....	72
Proficiency Effects on Textually Explicit Reading Questions	75
Proficiency Effects on Inferential Reading Questions	78
5. DISCUSSION.....	83
Findings on Treatment Effects.....	83
Findings on Proficiency Effects.....	89
6. CONCLUSION	101
Summary of the Findings.....	101
Limitations	104
Suggestions for Future Research.....	107
Final Conclusions.....	108
REFERENCES CITED.....	109
APPENDICES	
A. TAILORED CLOZE TESTS	114
B. SHORT-ANSWER TESTS.....	116
C. THE COURSE SYLLABUS OF ENGLISH II	124
D. TEXTS AND DISTRIBUTED WORKSHEETS FOR GROUPS 1 AND 2.....	127

LIST OF TABLES

Table	Page
1. Lower-level and Higher-level Processes Activated during Reading.....	10
2. A Comprehensive Taxonomy of Types and Contents of Reading Questions	15
3. A Taxonomy of Inferences	28
4. English Proficiency Levels of the Participants	44
5. Descriptive Statistics for Cloze Tests and Course Grades for Each Group of Participants	56
6. Descriptive Statistics for Textually Explicit and Inferential Reading Questions on the Pretest and Posttest for Groups 1 and 2	59
7. Levene's Test of Equality of Error Variances	59
8. Repeated-Measures ANOVA for Textually Explicit and Inferential Reading Questions	61
9. Descriptive Statistics for Textually Explicit Reading Questions on the Pretest and Posttest for Groups 1 and 2	63
10. Levene's Test of Equality of Error Variances	64
11. Repeated-Measures ANOVA for Textually Explicit Reading Questions	66
12. Descriptive Statistics for Inferential Reading Questions on the Pretest and Posttest for Groups 1 and 2	68
13. Levene's Test of Equality of Error Variances	68
14. Repeated-Measures ANOVA for Inferential Reading Questions.....	70
15. Descriptive Statistics for Textually Explicit and Inferential Reading Questions on the Pretest and Posttest for Three Proficiency Levels in Groups 1 and 2.....	73
16. The Estimated Marginal Means for the Main Effect of Proficiency	74
17. Bonferroni Test of Multiple Comparisons of Proficiency Levels	74

18. Descriptive Statistics for Textually Explicit Reading Questions on the Pretest and Posttest for Three Proficiency Levels in Groups 1 and 2	76
19. Descriptive Statistics for Inferential Reading Questions on the Pretest and Posttest for Three Proficiency Levels in Groups 1 and 2	79
20. The Estimated Marginal Means for Proficiency Levels	81
21. Bonferroni Test of Multiple Comparisons of Proficiency Levels	81

LIST OF FIGURES

Figure	Page
1. The relationship between test administration time and treatment group for textually explicit and inferential questions.....	62
2. The relationship between the test administration time and treatment group for textually explicit reading questions.....	67
3. The relationship between test administration time and treatment group for inferential reading questions	71
4. The relationship between group and proficiency for textually explicit and inferential reading questions	75
5. The relationship between treatment group and proficiency for textually explicit reading questions	78
6. The interaction between treatment group and proficiency for inferential reading questions	82

CHAPTER 1

INTRODUCTION

The Background of the Issue

People read to get something from text whether it be facts, ideas, enjoyment, or feelings (Nuttall, 2005). For whatever purpose they read, they naturally try to search for meaning in what they read (Graesser, Singer, & Trabasso, 1994; Nuttall, 2005), and meaning is constructed through readers' active cognitive interaction with the text (Nuttall, 2005; Tenma, Araki, Fukaya, Horiba, Isahara, Konishi, Miyaura, Murasugi, Narita, Ozaki, & Tajika, 2002). This active involvement of readers in the construction of meaning is reflected in their use of not only linguistic knowledge, but also background knowledge, world knowledge, personal experiences, and inferencing during reading (Grabe & Stoller, 2002)

As meaning is not fixed in the text, but is created in the active interaction of readers with the text, the readers' contribution to the construction of meaning is a key aspect of reading comprehension. In other words, the end product of reading comprehension is in part the result of linguistic and conceptual knowledge as well as the experiences that readers have accumulated in their life. Therefore, reading comprehension is, in a sense, a process in which readers project their life onto what they read (Tenma, 1992) and elaborate text information based on their background knowledge and experiences, both of which allow them to generate

knowledge-based inferences (Graesser & Kreuz, 1993). However, because experiences vary from person to person, the understanding of the same text can differ from reader to reader. This is particularly the case with the meanings that readers produce when they read in order to gain more than a literal understanding of the text information.

However, when it comes to reading activities in foreign language classrooms in Japan, too much emphasis is still placed on linguistic aspects of the text and on a fixed and analytical interpretation of words, phrases, or sentences at the cost of readers' active contribution to building up meaning. In this process of comprehension, information flows primarily one-way from the text into the mind of the readers, as the bottom-up processing of text information is emphasized. The result of this approach to reading is that different readers arrive at more or less the same conclusion concerning text meaning. This approach to reading is far from the interactive way that skilled readers use when creating meaning.

One way in which to alter this heavily bottom-up approach to teaching is through the effective use of reading questions, as they can provide a good alternative to the traditional grammar translation method in which readers depend excessively on bottom-up processing. Reading questions can be manipulated so that they encourage students to read the way that teachers want them to (Chikalanga, 1991; Kadota & Noro, 2001; Kanatani, 1995); hence, reading questions aimed at stimulating readers to read for more than a literal understanding can lead the readers to make use of their background knowledge,

and to engage in greater amounts of inferencing (Day & Park, 2005) as they depend more on top-down processing for creating meaning. Reading questions can help foster active involvement with the text and encourage the readers to read for meaning independently and in an interactive way.

Statement of the Problem

Teachers' Approaches to Reading Instruction

The first problem addressed in this study involves the pedagogical approaches used in much of the reading instruction in senior high school English classes in Japan. Reading instruction has predominantly been centered around the development of students' lower-level processing ability, with a particular emphasis on lexical access and syntactic parsing engendered by the use of Japanese translation ("Developing oral communication," 2007; Monbusho, 1999; Saito, 1992; JACET SLA research group, 2005). This traditional grammar translation method is based on the strong and in some cases blind belief among teachers that mastering language elements such as grammar and vocabulary must be prioritized if the students are to gain a high level of text comprehension. This situation, in which little attention is paid to higher-level processing involving the use of background knowledge and inferencing, is a typical example of the kind of foreign language classroom identified by Nuttall (2005) when she noted, "Many foreign language learners need to read for meaning. Language is simply the vehicle

conveying the message which is important to them, whereas in language classrooms the message is too often the vehicle conveying the language” (p. 30).

Terauchi (1991) surveyed Japanese teachers’ approaches to reading and revealed how reading instruction was implemented and the reasons for using that approach with 312 Japanese teachers of English in 104 high schools in the Tokyo area. Of the teachers surveyed, 74.3% answered that they invariably asked their students to provide a Japanese translation. Of these translation-minded teachers, 50.5% encouraged their students to copy the teachers’ translation in their notebooks. Some of the reasons that they gave for applying the traditional method were to ensure that the students will have a good grasp of the grammatical structures in the text and a detailed understanding of the text. Grammar translation was also seen as a way to help teachers deal with the decreasing English proficiency of their students and use an easy-to-apply teaching method. Finally, some teachers admitted that grammar translation was the only teaching method that they had experienced as a student and knew how to use as a teacher. With regard to teachers’ ignorance of other teaching methods, Kawasaki (1998) suggested that some teachers are unaware of pedagogical alternatives that have emerged in the last three decades of reading research. Furthermore, Mulvey (1998) proposed that teachers’ lack of familiarity with a variety of reading methodologies could underlie the perpetuation of current reading practices that place an excessive emphasis on bottom-up processing.

Japanese Students' Attitudes towards Reading

The second problem addressed in this study concerns Japanese students' attitudes towards reading. This problem is so closely related to the first problem that these two issues can be viewed as two sides of the same coin. Many Japanese students tend to read not for meaning but for a literal Japanese translation. Tsuchiya (1992) stated that such students are like slaves to literal translation who are likely to be always anxious about and addicted to reading intensively for translation. Therefore, in the process of arriving at the translation of the text, Japanese students read as decoders who linearly and discretely process words, phrases, and adjacent sentences (Monbusho, 1999; JACET SLA research group, 2005; Tenma et al., 2002) without referring to cognitive resources such as relevant background knowledge, world knowledge, experiences, or even context. The end product is inevitably the mechanical combination of conceptual meanings that words and phrases can have on their own, or, at best, the propositional meaning that a clause or sentence can have out of context. One commonly observed phenomenon in an English learning classroom is that even though the students have managed to puzzle out a Japanese translation of the text, when they are asked to explain what the text says (the contextual meaning) or to summarize what it is about (the pragmatic meaning), they are often unable to do so. They seem to have stopped processing textual meaning after managing to produce a word-for-word translation that allows them to comprehend only local and surface-level information in the passage. Consequently, they often view reading as a mechanical and laborious task,

far removed from the type of reading involving negotiation with text for constructing meaning through the active use of their linguistic and conceptual knowledge.

A survey conducted by Mulvey (1998) clarified Japanese students' passive attitudes toward reading and suggested why many of them form negative views of reading. He investigated 312 Japanese national university freshmen students' views of reading over two years and found that 72% of them identified reading activities that they had done in their EFL lessons with translation exercises. Mulvey (1999) added that as many as 92% answered that they had not been given any opportunities to discuss what they had read or to appreciate the texts from aesthetic or rhetorical perspectives. Instead, they had been required to memorize teachers' model translations as a means to preparing for taking tests.

In this way, many Japanese students read linearly as well as locally without integrating information from different parts of the text, inferring logical and conceptual gaps, or predicting. Rarely do they make use of their background knowledge in order to interrogate, criticize, or appreciate what the text says. As a result, Japanese students' approach to reading is far from a cognitive negotiation or interrogation with the text to create meaning. Students using this approach often cannot see the forest for the trees; they cannot see the whole message intended by the writer because of their short-sighted focus on local and surface information in the text.

Purposes of the Study

The first purpose of this study is to compare the effects of two types of reading questions on two comparison groups. Reading questions were divided into two types according to the level of comprehension that they ask for: textually explicit questions that require readers to read for a literal understanding of text, that is, to read locally for the propositional meaning of sentences and inferential questions that require readers to read for more than a literal understanding of text, that is, to read more globally for the contextual or pragmatic meaning of sentences. The second purpose of the study is to compare the effects of the two types of reading questions on students with differing English proficiency levels. For this purpose, the Japanese participants were divided into three proficiency subgroups (high, intermediate, and low), based on both the results of the tailored cloze test they took at the beginning of the experiment and their English course grades.

The Audience for the Study

This investigation into the effects of reading questions on text comprehension will potentially enable reading teachers to better understand what contribution inferential questions can make to helping readers comprehend text. An appropriate understanding of reading question types and their effects on text comprehension will help those teachers better match question types with the proficiency levels of students. At the same time, with this knowledge, teachers can make changes to their classroom practices and break the cycle of translation-bound

reading instruction with its overemphasis on lower-level processing. Accordingly, they will lead their students not to be decoders but to be meaning-makers of a text.

Delimitations

This study is empirical classroom research on reading instruction with homogeneous Japanese EFL learners. The Japanese participants are senior high school students aged from 16 to 17 whose English proficiency level is generally low in part because they seldom have chances to use English outside of their classrooms and in part because they learn English mainly in order to pass entrance exams into Japanese universities where English is usually one of the required subjects.

The Organization of the Study

Chapter 2 is a review of the literature on the mental processes of reading, the function and taxonomies of reading questions, and a taxonomy of inferences from empirical studies on reading questions and inferences, respectively. At the end of the chapter, two research questions are presented. In Chapter 3, the methodology of the study is described in terms of the participants, instrumentation, research design, and procedures. The results of the study are presented in Chapter 4. Chapter 5 is a discussion of the results of the study with answers to the research questions. Chapter 6 describes the conclusion of the study and suggestions for future studies.

CHAPTER 2

REVIEW OF THE LITERATURE

This review of the literature includes an overview of the mental processes involved in reading based on the view of Grabe and Stoller (2002), an exploration into the functions of reading questions, and an examination of the taxonomies of reading questions and inferences suggested by reading teachers and researchers, along with empirical studies of Japanese EFL readers and of reading questions used in the major tests in Japan.

What Processes Does Reading Involve?

One of the most urgent tasks facing Japanese English teachers and reading teachers in general (Birch, 2007) is understanding the processes that are involved in reading comprehension and then providing their students a clear and appropriate explanation of how proficient readers read. By so doing, teachers can help free the students from the fossilized blind belief that they must read word-by-word or sentence-by-sentence in order to obtain a propositional meaning from the surface-level information. At the same time, teachers may be able to change students' passive and negative attitudes toward reading (Tsuchiya, 1992).

What processes does reading involve? According to Grabe and Stoller (2002), there are two classes of common underlying processes that are activated in reading: lower-level and higher-level processes (see Table 1).

Table 1. *Lower-level and Higher-level Processes Activated during Reading* (Grabe & Stoller, 2002, p. 20)

Lower-level processes	Higher-level processes
Lexical access	Text model of comprehension
Syntactic parsing	Situation model of reader interpretation
Semantic proposition formation	Background knowledge use and inferencing
Working memory activation	Executive control processes

Lower-level processes concern linguistic processes that occur relatively automatically in proficient readers. When words are read, their meanings are accessed, and their grammatical functions are extracted via syntactic parsing. Then, based on the combination of the semantic information and the grammatically analyzed information, the most logical clause-level meaning units, that is, propositions (in the code of who did what) are formed (semantic proposition formation). These lower-level processes operate in working memory; once lower-level processing begins operating, working memory is accordingly activated. In this activated short-term memory, a network of information arising from each process is established. The network of information is stored in working memory until it is integrated into what has been read in order to make sense of the meaning encoded in the text.

The higher-level processes, which include comprehension and knowledge processes, critically involve reader's background knowledge. This combination

allows inferencing to take on an important role. A monitor, which is called executive control processing, is involved in the construction of both a text model of comprehension and a situation model of reader interpretation. The text model of reading comprehension, which is a mental representation of what the text says, is the most fundamental higher-level comprehension process. It is based entirely on the text information and “amounts to an internal summary of main ideas” (Grabe & Stoller, 2002, p. 26). In contrast to the text model, the situation model of reader interpretation is a reader’s actively constructed “microworld of agents, objects, actions, events, spatial composition, and states during comprehension. This mental microworld is analogous to everyday experiences in the physical and social world” (Graesser & Kreuz, 1993, p. 147). The situation model of reader interpretation is visually represented like a cartoon in readers’ minds (Tenma et al., 2002). This reader interpretation is built on the text model but is elaborated on the basis of the reader’s experiences, background knowledge, and knowledge-based inferences (Graesser & Kreuz, 1993). In this way, reader interpretations deeply involve the nature and mental state of readers. Thus, different readers’ interpretations of the same text often differ even though the interpretations are based on the same text information. Background knowledge and inferencing make a more significant contribution to transforming clause-level meaning units into the text model of comprehension and further into the elaborated situation model of reader interpretation than lower-level processes. Once the reader interpretation is created, it is accommodated into long-term memory. As a result of revising and deepening,

which involves transforming knowledge in the process of storing it in long-term memory, readers learn from reading (Heilman, Blair, & Rupley, 1986; Kadota & Noro, 2001; Nuttall, 2005).

An appropriate understanding of these processes involving the combination and integration of various sources of knowledge including both lower-level and higher-level knowledge sources that occur during skilled reading can provide teachers with a grasp of the role that students need to play when reading (Birch, 2007). With this knowledge of proficient reading processes, teachers should explicitly encourage readers' active involvement in terms of both linguistic and cognitive interactions with the text. In this way, grammar-translation reading lessons, which are excessively dependent on lower-level processing, can be transformed into reading classes in which background knowledge and inferencing receive a greater emphasis. This transformation will also help readers strike a balance between lower-level and higher-level processing by facilitating interactive processing and promoting students' active contribution to building, interpreting, or creating text meaning.

Functions of Reading Questions

Task-based reading instruction is one pedagogical alternative to the grammar-translation method. Reading tasks can be manipulated so that they encourage students to read the way that teachers want them to (Chikalanga, 1991; Kadota & Noro, 2001; Kanatani, 1995). For example, tasks that activate students'

use of background knowledge and inferences can lead them to interact with the text and thus to create a personalized interpretation of text meaning (Day & Park, 2005).

One common type of reading task involves the use of questions. Reading questions have several functions. Their first function is to encourage students to engage in reading for meaning (Heilman, Blair, & Rupley, 1986; Nuttall, 2005), which is essentially what people naturally do during reading (Graesser, Singer, & Trabasso, 1994; Nuttall, 2005). Reading questions can also motivate students to read because they will read purposefully in order to find the answers (Grant, 1987; Heilman, et al., 1986; Nuttall, 2005). Although readers approach texts in different ways according to their reading purposes, some purpose or task imposed either by the readers themselves or the teachers can act to summon up motivation to read (Grabe & Stoller, 2002). Second, reading questions can function to direct students' attention to the important aspects of the text (Nuttall, 2005; Vacca, Vacca, & Gove, 1991). For instance, signpost questions, which function similarly to a signpost at a crossroads that shows travelers the right road to take, can focus readers' attention on such global aspects of text as the theme of an article or topic sentences of its paragraphs. Finally, questions can encourage students to regard reading as a means to look for answers; this kind of attitude towards reading may help reduce the psychological burden students often feel when they are engaged in foreign language reading (Kanatani, 1995), and in particular Japanese students who often tend to view reading as a mechanical and laborious task. Unfortunately, many teachers are not well aware of these useful functions reading questions have for

readers, and the influence that they can exert on the level of thinking that occurs (Heilman, et al., 1986).

Taxonomies of Reading Questions

It is important for teachers to be aware of the type and content of questions that they ask students when they read (Heilman, et al., 1986; Cohen & Upton, 2006). Reading questions can be classified according to the degree of cognitive activity that they encourage, the type of reading skills that they require, the levels of comprehension that they promote, and the degree of reading proficiency that they demand.

Taxonomies, which categorize reading questions in the aforementioned ways, can be helpful for teachers by providing a framework for (a) structuring and evaluating questions (Day & Park, 2005; Heilman, et al., 1986; Nuttall, 2005), (b) developing their own questions (Day & Park, 2005; Ehara, 1996; Heilman, et al., 1986; Nuttall, 2005), and (c) balancing the types of questions that they ask students (Day & Park, 2005; Ehara, 1996; Nuttall, 2005). Table 2 shows a comprehensive taxonomy into which reading questions have been categorized by several authors and researchers. In the following sections, I will describe six categories of reading questions.

Table 2. *A Comprehensive Taxonomy of Types and Contents of Reading Questions*

Categories of Questions	Reading Skills	Levels of Comprehension	Reading Abilities	Question and Answer Relationships	Convergence and Divergence
Low-order questions	Type 1 (Nuttall, 2005) Plain sense reading questions (Grant, 1987) Literal (Day & Park, 2005)	Literal level (Heilman, et al., 1986; Vacca, et al., 1991)	Read the lines	Right there (Vacca, et al., 1991) Textually explicit (Pearson & Johnson, 1978)	Narrow (Vacca, et al., 1991) Convergent (Heilman, et al., 1987; Pearson & Johnson, 1978)
Higher-order questions	Types 2 and 3 (Nuttall, 2005) Deductive reading questions (Grant, 1987) Reorganization and Inference (Day & Park, 2005)	Interpretive level (Heilman, et al., 1986; Vacca, et al., 1991)	Read between the lines	Search and think (Vacca, et al., 1991) Textually implicit (Pearson & Johnson, 1978)	Close-ended Broad (Vacca, et al., 1991)
	Types 4, 5, and 6 (Nuttall, 2005) Projective reading questions (Grant, 1987) Prediction, Evaluation and Personal response (Day & Park, 2005)	Critical level (Heilman, et al., 1986) Applied level (Vacca, et al., 1991)	Read beyond the lines	On my own (Vacca, et al., 1991) Scriptally implicit (Pearson & Johnson, 1978)	Divergent (Heilman, et al., 1987; Pearson & Johnson, 1978) Open-ended

Low-order and Higher-order Questions

The first taxonomy of reading questions in the left column of Table 2 is titled Categories of Questions. It shows that reading questions can be broadly categorized into two subgroups according to the required level of cognitive operation: lower-order questions and higher-order questions. The lower-order questions demand the recognition or recall of factual information explicitly presented in the text. The explicitly stated text information generally involves facts, names, dates, times, locations, lexical items, and propositions. In contrast, higher-order questions require more than recognition or recalling verbatim information directly accessed from the text. They also facilitate moving beyond a literal understanding of the text to a more knowledge-based and global understanding of textual meaning. In other words, they require readers to read between the lines or beyond the lines.

Types of Reading Questions Based on Reading Skills

The second class of reading questions in the second column of Table 2 is titled Reading Skills. These questions are classified according to the skills that they require from readers. Thus, this is a reader-bound part of the taxonomy. Nuttall (2005) presented a taxonomy of questions that is made up of six categories, and explained those types of questions with a sample passage and accompanying questions illustrating each related type. The six categories are as follows:

Type 1: Literal comprehension questions. Type 1 questions are considered the same as the lower-order questions. The answers to Type 1 questions can be often answered in the words of the text.

Type 2: Reorganization or reinterpretation questions. Type 2 questions require readers either to reinterpret the literal information found in the text or to obtain bits and pieces of surface information from different parts of the text and to put them together in a new way. Elementary inferencing is sometimes involved in arriving at the answer.

Type 3: Inferential questions. Type 3 questions oblige readers to consider what is implied but not explicitly stated in the text. Type 3 questions are regarded as more cognitively challenging than the first two question types because readers are required to understand the text well enough to make logical and conceptual inferences.

Type 4: Evaluative questions. Type 4 questions demand that readers make a considered judgment about the text in terms of what the writer is trying to do and how well he or she has achieved it.

Type 5: Personal response questions. Type 5 questions require readers to make a personal reaction to the text, based on an appropriate understanding of the text.

Type 6: Questions concerned with how writers say what they mean. Type 6 questions ask for the readers' opinion about rhetorical values of the text, i.e., the writer's way of expressing ideas and organizing the text.

Nuttall (2005) found that questions in foreign-language textbooks tended to be mainly type 1 questions with a few type 2 and 5 questions. As a result, she suggested that in addition to questions that ask for a literal understanding of and personal responses to what writers say in a text, questions that ask about how and why writers write a text should be employed more frequently. Furthermore, she stated that even evaluative questions can often be answered by beginning learners using a largely top-down approach.

Grant (1987) listed three basic reading skills and provided samples of three question types that can be used to activate those skills. His question taxonomy consisted of (a) plain sense reading, (b) deductive reading, and (c) projective reading. Plain sense reading, which is the most basic reading skill, corresponds to the ability to understand the plain sense of what is stated in the text. In other words it is the ability to comprehend information explicitly stated in the text. Deductive reading matches the ability to draw inferences from the text information, that is, the ability to read between the lines. Projective reading, which is the most advanced category of reading skills, amounts to the ability to read beyond the lines. This skill involves the ability to relate the reading passage to real life; in particular to the reader's own opinions, knowledge, imagination and experience. He proposed that both deductive and projective reading skills should be a part of a reading curriculum even at the beginners' level.

Day and Park (2005) emphasized the interactive process in which readers construct meaning and proposed a taxonomy of six types of comprehension

questions, based on the studies of Pearson and Johnson (1978) and Nuttall (2005): (a) literal comprehension, (b) reorganization, (c) inference, (d) prediction, (e) evaluation, and (f) personal response. The definitions of all these categories, except for prediction, correspond roughly to those of Nuttall (2005). Prediction questions involve readers in using not only their understanding of the passage but also their own knowledge of the topic and related matters in a systematic fashion so as to determine what might happen next during reading or after a story ends.

Day and Park's research on reading and their observations of reading classrooms demonstrated that effective teachers and teachers in effective schools ask more higher-order questions that force readers to go beyond a literal understanding of a text. They also found that students tended to perform well on the types of comprehension questions that had been repeatedly employed by their teachers. For these reasons, they suggested that teachers should teach students how to go beyond a literal level of understanding and provide them opportunities to engage with all six types of comprehension questions. However, they did not clarify which proficiency level of students should be given which types of comprehension questions, and to what degree students will be able to develop their reading abilities if an emphasis is placed on higher-order questions.

*Types of Reading Questions Based on Levels of Comprehension and Reading
Abilities*

Different forms of reader-bound taxonomies have been established on the basis of the levels of comprehension at which readers respond to meaning (see the third column in Table 2, Levels of Comprehension), and on the abilities that they demand of readers (see the fourth column, Reading Abilities).

Heilman, Blair, and Rupley (1986) suggested that properly framing classroom reading questions in relation to desired reading comprehension outcomes is necessary if teachers are to develop higher levels of purposeful reading in their students. Providing students with specific prereading questions helps engage them in learning, direct their purposes for reading, and regulate their depth and rate of reading. Based on their belief that taxonomies should be used as frameworks to structure the use of comprehension questions that will cover a wide variety of levels and abilities, they proposed three levels of comprehension: (a) literal, (b) interpretive, and (c) critical comprehension. Literal comprehension abilities include word recognition, the recall of details and main ideas directly stated in a text, syntactic processing, and knowledge of passage structure. Interpretive comprehension abilities involve the abilities to reason from the text information in order to determine the author's tone, purpose, and attitude, to infer factual information, main ideas, comparisons, and cause-effect relationships not explicitly stated in the passage, and to summarize story content. Critical comprehension abilities encompass analyzing and evaluating the quality of text information in

terms of some standard, and, on the basis of the analysis and evaluation, to react personally to the information in the text.

Vacca, Vacca, and Gove (1991), based on the concept that “readers respond to meaning at various levels of abstraction and conceptual difficulty” (p. 163), asserted that reading teachers can engage children in a full range of cognitive functions by using levels of comprehension to plan and guide instructional activities. Similar to Heilman, et al.’s (1986) classification mentioned above, Vacca, et al. presented three levels of comprehension: (a) literal level, (b) interpretive level, and (c) applied level.

Types of Reading Questions Based on Question and Answer Relationships

Pearson and Johnson (1978), Raphael (1982, 1986), and Vacca et al. (1991), added a new perspective to the taxonomy of reading questions and classified reading questions based on the relationships between the questions and answers. Vacca, et al. (1991) proposed that, “The three-way relationships that exist among the question, the text to which it refers, and the background knowledge and information at the reader's disposal” (p. 165) can help specify which information sources readers can resort to when they look for answers to different types of questions. The taxonomy is presented in the second column from the right in Table 2, Question and Answer Relationships.

The first information source that readers tend to access when answering questions is the text, as answers are directly, explicitly, and locally stated in the text.

Therefore, readers can answer these questions using the words found in the text (Nuttall, 2005). These text-bound questions have been labeled Textually Explicit Questions (Pearson & Johnson, 1978) or Right There (Raphael, 1982, 1986) because the words used in the question and in the most appropriate answer are “right there” in a single sentence (Raphael, 1982, p. 187).

The second information source readers resort to in answering questions is also the text, but the answers are not as obviously presented in the text as the answers to textually explicit questions. At least one step of logical or pragmatic inferencing is needed in the process of arriving at the answer. Readers are required to search more globally for explicitly stated information in various parts of the text and to infer the hidden relationships among them. Questions that are more knowledge-based than text-bound questions and whose answers are found on the page but are not explicitly expressed are labeled Textually Implicit Questions (Pearson & Johnson, 1978) or Search-and-Think (Raphael, 1982, 1986).

The other information source is the reader. Because the answers to questions that are based on reader knowledge come primarily from the readers’ background knowledge, inferencing is an important component of this answering process. Inferencing activated by knowledge-based questions differs from inferencing facilitated by textually implicit questions in terms of the information source that is accessed for inferencing. Inferencing in knowledge-based questions depends primarily on the readers’ knowledge and inferencing in textually implicit questions depends primarily on information on the page. These knowledge-based

questions, whose answers largely reside in the readers' minds, are labeled Scriptally Implicit Questions (Pearson & Johnson, 1978) or On My Own (Raphael, 1982, 1986).

Convergence and Divergence Reading Questions

Vacca et al. (1991) also classified reading questions as narrow and broad, based on Cunningham's (1971) view of question-asking relationships along a continuum from narrow to broad. The taxonomy based on this view is shown in the right column of Table 2, Convergence and Divergence.

The narrowest questions on the continuum ask readers to prompt literal understanding of text information. Often this includes the ability to recognize information that answers *who*, *where*, *what*, *when*, or *how many* questions. Narrow (Vacca et al., 1991) and convergent (Heilman, et al., 1986; Pearson & Johnson, 1978) questions engage readers in more complex cognitive activity at an interpretive level. In order to draw inferences, readers must converge or narrow the search down on text information and identify textual relationships such as cause-effect relations, as well as similarities and differences among elements in the text. Readers must also use information from different parts of the text to interpret the author's intended meaning, especially in narrative or expository texts. Because of the low possibility of varying answers and the conventional rhetorical organization in the content of narrative or expository texts, it is likely that readers will read in a convergent way. Consequently, the answers that readers provide to those questions

can be predicted because the answers can be objectively classified as right or wrong; thus, they are also known by the teachers in advance.

In contrast, broad (Vacca et al., 1991) and divergent (Heilman, et al., 1986; Pearson & Johnson, 1978) questions require readers to understand information at more than a literal level. As they lead to divergent thinking, readers' responses are unpredictable, the answers cannot be categorized right or wrong, and they are not known to teachers in advance. In this regard, Vacca et al. (1991) stated, "Narrow questions prompt literal understanding and inferential thinking. Broad questions lead to applied responses and engage readers in creative and evaluative thinking" (p. 164). In sum, the narrower the question becomes, the more convergent the answer becomes, whereas the broader the question, the more divergent the responses.

Empirical Studies on Japanese EFL Readers and Reading Test Questions

In this section I will look at two empirical studies on reading. The first study was an investigation into what information sources skilled and less skilled Japanese EFL readers depended on in comprehending text. In the second study, the major English proficiency tests in Japan were surveyed and examined in terms of what type of reading questions were employed in them.

Yamashita (2003) investigated what information sources skilled and less skilled readers depended on in answering a gap-filling test in order to examine whether such tests can measure textual comprehension. Twelve Japanese university EFL students were divided into two groups of skilled readers and less skilled

readers according to reading comprehension test scores. All the participants completed a 16-item gap-filling test while thinking aloud about their test-taking processes. She found that the more skilled readers tended to employ text-based information both locally and globally, while the less skilled readers were more likely to use text-based information locally at the clause or sentence level and to spend more time decoding the text information.

Shimizu (2005) examined what types of reading questions were employed in three major English tests in Japan: the Test of English as a Foreign Language (TOEFL), the Society for Testing English Proficiency (STEP), and the University Testing Center (UTC) tests. She classified the questions into six categories: paraphrasing (straightforward wh-questions that require readers to find local explicit information), inference, thematic, referential, vocabulary, and text organization. Based on Graesser et al.'s (1994) classification scheme of inferences, she subcategorized inference questions into causal antecedent, superordinate goal, character emotional reaction, causal consequence, instrument, subordinate goal-action, state, and emotion of reader. She found that the question types used in both the STEP and TUC tests were limited to only two types of questions, paraphrasing and inference questions (causal antecedent, superordinate goal-action, and character emotional reaction), in comparison with the TOEFL test, which involved almost all types of reading questions in a more-balanced fashion.

Taxonomies of Inferences Involved in Reading

“Text is never totally explicit” (Chikalanga, 1992, p. 697) because writers do not typically spell out facts and relationships that they think that readers are aware of (Nuttall, 2005). In order to make sense of a text, readers need to fill in conceptual gaps locally within or between sentences and globally between paragraphs. The ability to coherently reconstruct a writer’s unstated presuppositions by connecting text information and background knowledge requires inferencing. This ability plays an indispensable part in text comprehension. If readers cannot make use of this ability while reading, they will have difficulty making meaning representations of text (Grabe & Stoller, 2002). Without readers’ attempts to reconstruct the underlying presuppositions, text is no more than a collection of words and paragraphs (Kembo, 2001) that cannot stand on their own (Tenma et al., 2002).

The generation of inferences can be facilitated by particular tasks (Graesser et al, 1994; Kadota & Noro, 2001). Chikalanga (1992) stated that if reading teachers first understand types of inferences and the levels of cognitive operation that they encourage, and then develop questions and tasks based on that understanding, they can help their students to achieve higher levels of comprehension. In addition, a final goal may be that the students are able to make good use of inferences while reading by themselves. Then, what kinds of inferences are believed to be involved in text comprehension and how do they function?

Inferences have been categorized generally in two ways: (a) what information sources readers depend on for generating the inferences (text-based information or non-text based information) and (b) when the inferences are generated (during the course of comprehension, after text comprehension, or during a later task). In the next section, I will discuss the information sources on which readers base their inferences.

A Taxonomy of Inferences Based on Information Source

Chikalanga (1992, 1993) suggested a comprehensive taxonomy of inferences generated by inferential questions in order to provide teachers with a system for constructing various types of reading questions that can be used in reading instruction and for assessing questions used in teaching materials. She proposed that inference is central to reading comprehension because writers do not explicitly state what they think that readers already know. As a consequence, readers are inevitably required to fill in conceptual gaps in order to comprehend the author's intended meaning.

First, as a broad theoretical framework for classification, she focused on information sources that readers resort to when making inferences. Chikalanga based her classification on Pearson and Johnson's (1978) taxonomy of reading questions where inferences are divided into text-based (textually implicit) and schema-based (scriptally implicit) ones. Here, she made a clear distinction between inferences that are constrained by the propositional content of the text and those

that are constrained by the prior knowledge of the reader. Next, on the basis of Warren, Nicholas, and Trabasso's (1979) taxonomy and Nicholas and Trabasso's (1980) taxonomy, Chikalanga further classified inferences into three basic categories: lexical, propositional, and pragmatic (or scriptal) inferences (see Table 3).

Table 3. *A Taxonomy of Inferences* (Chikalanga, 1992, p. 705)

Basic Category	Specific Types	Question-Answer Relationship
Lexical	(a) Pronominal inferences (b) Ambiguous/unfamiliar word meanings	Textually/Scriptally Implicit
Propositional	(a) Logical Informational referential spatio-temporal (b) Logical Explanatory motivational causative enablement	Textually Implicit
Pragmatic/ Scriptal	(a) Elaborative Informational referential spatio-temporal (b) Elaborative Explanatory motivational causative enablement (c) Evaluative	Scriptally Implicit

Lexical inferences are defined as inferences that readers make to identify the referents of pronominals and to disambiguate unknown lexical terms from the context in which they occur. Readers retrieve the referent from the text, the context, and/or their prior knowledge (Nuttall, 2005). As lexical inferences are generated either on the text information (textual-bound) or on readers' schemata (reader-bound), they are under the category of textually or scriptally implicit in Pearson and Johnson's (1978) taxonomy.

Propositional inferences are defined as inferences that readers draw logically based on the semantic content of explicitly stated propositions in the text. There are two types of propositional inferences in this category: (a) logical informational inferences and (b) logical explanatory inferences. Logical informational inferences are made in response to referential questions, (i.e., wh-questions such as who and what) and spatio-temporal questions (i.e., wh-questions such as where and when). Logical explanatory inferences are made in response to motivational and causative questions and questions about conditions that enable actions or events to occur (i.e., questions beginning with why and how). Propositional inferences are categorized as textually implicit inferences in Pearson and Johnson's (1978) taxonomy because they are all text-bound inferences.

Pragmatic inferences are defined as inferences that readers generate depending largely on background knowledge. There are three types of pragmatic inferences in this category: informational, explanatory, and evaluative inferences. The first two types are the same as those in the basic category of propositional

inferences, but the information source is within the readers, not the text. Chikalanga added the word *elaborative* to these two types in order to distinguish text-bound inferences from their reader-bound counterparts. Evaluative inferences are drawn to “assess the significance, normality, morality and validity of events, characters’ actions, etc. specified within the text” (Chikalanga, 1992, p. 702). Pragmatic inferences are strongly dependent on the readers’ knowledge and on their belief or moral systems. They are categorized as scriptally implicit inferences in Pearson and Johnson’s (1978) taxonomy because they are all knowledge-based inferences.

Chikalanga demonstrated the workability of her taxonomy using the following short passage with questions, plausible answers to them, and the types of inferences that would be generated in order to answer the questions.

*It was Friday afternoon. Carol was drawing a picture in the classroom.
David felt mischievous. David decided to tease Carol. When Carol was
not looking, he tied her shoelaces together. Carol tripped and fell down.*

1. Question: Why did David decide to tease Carol?

Response: Because of his mental state of feeling mischievous.

Inference Type: Logical Explanatory (motivational) inference.

(Note that the motivation for David’s decision to tease Carol is specified in the text but it is not linguistically marked, so it is implicit.)

2. Question: Why, do you think, was Carol drawing a picture?

Response: Because artwork gave her pleasure or satisfaction and this perhaps explains why she was not aware of what David was trying to do to

her.

Inference Type: Elaborative Explanatory (motivational) inference.

(Note that, in this case, the motivation for Carol's action is not specified in the text. The reader has to provide it based on their experience or prior knowledge.)

3. Question: What instrument was Carol using in drawing the picture?

Response: A pencil, crayon, or pen (because the predicate is 'draw,' not 'paint').

Inference Type: Elaborative Informational (referential).

4. Question: Why did Carol trip?

Response: Because her shoelaces were tied together.

Inference Type: Logical Explanatory (causative).

5. Question: Why was it possible, do you think, for David to tie the shoelaces together without being noticed?

Response: Because Carol was too absorbed to notice what David was doing.

Inference Type: Elaborative Explanatory (enablement) inference.

(Chikalanga originally categorized this inference as Logical Explanatory (enablement) inference, but Carol's state of mind is not specified in the text when her shoelaces were being tied, so the reader must fill in this pragmatic gap logically by referring to world knowledge in the context where 'she was not looking.')

6. Question: What were Carol and David, do you think?

Response: Probably school children.

Inference Type: Elaborative Informational (referential) inference.

7. Question: What kind of shoes was Carol wearing?

Response: Lace-up shoes.

Inference Type: Logical Informational (referential) inference.

8. Question: Where did this incident happen?

Response: At school (we are told Carol was drawing a picture in the classroom).

Inference Type: Logical Informational (spatial) inference.

9. Question: Who does 'he' refer to?

Response: David.

Inference Type: Lexical (pronominal) inference

10. Question: Was David mean?

Response: Perhaps not, assuming that the two were friends and that they probably teased each other from time to time

Inference Type: Evaluative. (Chikalanga, 1992, pp. 705-707)

A Taxonomy of Inferences Based on Generation Time

It is also possible to categorize inferences in terms of discourse processing by considering the time at which inference generation occurs. This concerns whether the inferences are made while comprehending the text or during a later task (Graesser & Kreuz, 1993; Graesser, Singer, & Trabasso, 1994). Graesser,

Singer, and Trabasso (1994) proposed thirteen types of knowledge-based inferences for text comprehension based on the constructionists' principle of "search after meaning" (p. 377). This principle asserts that when readers comprehend narrative text, they actively attempt to construct meaning representations (in the form of a text model of reading comprehension and a situation model of reader interpretation) (a) that are coherent at local and global levels among events, actions, and states in the text; (b) that explain explicit information such as actions, events, and states mentioned in the text; and (c) that satisfies the readers' goals. The thirteen classes of inferences are described briefly as follows:

Class 1: Referential

A word or phrase is referentially tied to a previous element or constituent, explicit or inferred, in the text. For example, in "...on removing the *fork* the eye came with *it*," *fork* is the referent of *it*.

Class 2: Case structure role assignment

The inference is on grammatical analysis by which an explicit noun phrase is assigned to a particular case structure role, e.g. agent, recipient, object, location, or time.

Class 3: Causal antecedent

The inference is on a causal chain or bridge between the current explicit action, event, or state and the previous message context.

Class 4: Superordinate goal

The inference is a goal that motivates an agent's intentional action.

Class 5: Thematic

This is a main point or moral of the text.

Class 6: Character emotional reaction

The inference is an emotion experienced by a character, caused by or in response to an event or action.

Class 7: Causal consequence

The inference is on a forecasted causal chain, including physical events and new plans of agents. These inferences do not include the character emotions in class 6.

Class 8: Instantiations 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. For example, breakfast is instantiated by egg and bacon.

Class 9: Instrument

The inference is an object, part of the body, or resource used when an agent executes an intentional action.

Class 10: Subordinate goal-action

The inference is a goal, plan, or action that specifies how an agent's action is achieved.

Class 11: State

The inference is an ongoing state, from the time frame of the text, which 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.

Class 12: Emotion of the reader

The inference is the emotion that the reader experiences when reading a text.

Class 13: Author's intent

The inference is the author's attitude or motive in writing. (Graesser, Singer, & Trabasso, 1994, p.375)

Graesser et al. (1994) defined the thirteen classes of inferences according to the content of the inferences and their relation to explicit information in the text. These classes of inferences are broadly divided into three categories according to when they are generated: on-line (during the course of comprehension), off-line (after text comprehension during a later task), and unidentified (on-line and/or off-line).

Classes 1 through 6 are considered to be generated on-line, and these on-line inferences are called bridging inferences or backward inferences because the information being comprehended is related to previous explicit information in the text (Kadota & Noro, 2001; Tenma et al., 2002).

Referential (Class 1), case structure role assignment (Class 2), and causal antecedent (Class 3) inferences occur during the establishment of local coherence,

which is achieved when conceptual connections are made between the contents of adjacent text constituents like a phrase, proposition, and clause or between short sequences of constituents. In other words, readers establish local coherence by filling in logical and pragmatic gaps within a sentence or between adjoining sentences.

Superordinate goal (Class 4), thematic (Class 5), and character emotional reaction (Class 6) inferences are essential for establishing global coherence, which is established when local chunks of information are interconnected and organized into higher-order chunks. In other words, readers achieve global coherence by filling in logical and conceptual gaps among more than adjacent sentences or between paragraphs.

Causal consequence (Classes 7), instantiation of noun category (Class 8), instrument (Class 9), subordinate goal-action (Class 10), and state (Class 11) inferences are considered to be constructed off-line. Unlike the on-line inferences, these off-line inferences are not required in order to establish coherent explanatory meaning representations (Graesser et al, 1994), but they are helpful in deepening the readers' understanding of the text (Kadota & Noro, 2001; Tenma et al., 2002). They are called elaborative inferences or forward inferences because readers elaborate on the information comprehended or predict what will happen next depending on textual information and/or reader-bound information.

Emotion of reader (Classes 12) and author's intent (Class13) "address the pragmatic communicative exchange between reader and author" (Graesser et al., 1994, p. 376), but it is unclear when they are constructed.

Tajika (1999) proposed that Class 1, Class 2, and Class 3 inferences are mainly based on explicit text information, whereas the remaining inferences are based on individual readers' conceptual and experiential knowledge about the world. Thus, the text-based inferences (Class 1 through Class 3) are essential elements in constructing a proper understanding of a text.

Tajika (1999) investigated whether EFL readers generated causal-consequence inferences when reading a narrative passage, and how those inferences affected text comprehension. Two groups of participants took part in the study: thirty Japanese college juniors whose majors were related to the English language and culture or a related field and eighteen Japanese university freshmen who were not majoring in English. All of the participants read an English passage that consisted of seven sentences with three reading questions that were designed to (a) check comprehension of a key sentence, (b) elicit a causal antecedent inference (a bridging inference), and (c) elicit a causal consequence inference (an elaborative inference). The participants were instructed to read the text sentence by sentence. While they were reading, they were asked to write in Japanese whatever idea entered their mind in response to the designed three questions.

Tajika found that causal-consequence inferences may not necessarily be encoded on-line by every reader. More importantly, she discovered that about half

of the EFL readers (fifteen out of thirty college juniors with successful comprehension of the key sentence, and five out of eleven university freshmen with successful comprehension) made no inferences while reading; they read in a “comprehension only” reading style (p. 141). Based on this finding, she suggested that teachers should promote more inferential reading in which learners try to make conceptual connections between textual information and fill in logical and pragmatic gaps in the text.

Summary of Literature Review

Taxonomies of reading questions have been established largely as parts of approaches to reading instruction (Grant, 1987; Heilman et al., 1986; Pearson & Johnson, 1978; Vacca, et al., 1991; Raphael, 1982, 1986; Takanashi & Takahashi, 1987; Thomas, 1992), and as theoretical frameworks for research on classroom questioning (Day & Park, 2005), and on comprehension questions employed either in textbooks (Applegate, et al., 2002; Nuttall, 2005) or used on tests (Champeau De Lopez, Marchi, & Arreaza-Coye, 1997; Shimizu, 2005). In a similar way, taxonomies of inferences have been established as parts of approaches to reading instruction (Chikalanga, 1992) and as theoretical frameworks for research on inferences (Chikalanga, 1993; Graesser et al., 1994). They have also been used for research on the generation of inferences elicited by reading questions (Tajika, 1999).

Research on reading questions has revealed that questions used in textbooks, those asked by teachers in classrooms, and even those employed in major English proficiency tests in Japan tend to primarily ask for literal or text-bound comprehension (Applegate, Quinn, & Applegate, 2002; Ehara, 1996; Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt, & Schedl, 2000; Nuttall, 2005; Shimizu, 2005; Thomas, 1992). Research on inferences elicited by reading questions has also shown that quite a few Japanese university EFL learners are likely to focus on the literal comprehension level without making any inferences (Tajika, 1999). It seems that not only the reading materials and pedagogical classroom practices adopted by many Japanese teachers of English but also reading questions employed in major English tests conducted in Japan do not encourage learners to read beyond a literal understanding, or to engage in active cognitive interaction with the text by filling in logical and conceptual gaps with inferencing and their background knowledge. Japanese EFL learners' approach to reading provides strong evidence to support this claim and reflects the current pedagogical practices used in many English reading classrooms. Therefore, it is strongly suggested that more inferential questions that stimulate learners' conceptual and experiential knowledge should be utilized in order to encourage foreign language learners to positively interact with text in deeper and more meaningful ways.

However, little empirical research has been conducted on what effect reading instruction with an emphasis on inferential questions will have on reading

comprehension. Research into the effect of different types of reading questions on readers with varying proficiency levels is also lacking.

Research Questions

The purpose of this study is to determine which type of questions, textually explicit (TE) or inferential (IF) questions, will best facilitate text comprehension, and which type of questions has the most beneficial effect on Japanese EFL senior high school students at three proficiency levels (low, intermediate, and high). More specially, the following research questions will be investigated:

1. To what degree will the test scores differ when students receive instruction emphasizing IF questions or instruction emphasizing TE questions?

Hypothesis 1: Gains will be greater for the group that receive instruction emphasizing IF questions than for the group that receive instruction emphasizing TE questions because IF questions will encourage readers to fill in logical or pragmatic gaps during reading by making active use of background knowledge and inferences and thus, facilitate readers' cognitive involvement more with the text for meaning.

2. To what degree will the test scores differ for high, middle, and low proficiency participants?

Hypothesis 2: Gains will be the greatest for participants at the high proficiency level within the group that receive instruction emphasizing IF reading questions because IF questions will match best the way that proficient readers read a text

interactively.

Hypothesis 3: Gains will be the greatest for participants at the lowest proficiency level within the group that receive instruction emphasizing TE questions because TE questions will match best the way unskilled readers read a text, largely depending on bottom-up processing.

CHAPTER 3

METHODS

Participants

The participants in this study were Japanese public senior high school students, aged 16 or 17, studying in the same school. Although they had been studying English in an EFL setting for at least five years, they rarely had chances to use English outside of the classroom. The school is a highly competitive coeducational institution with almost an equal number of male and female students. Almost all the students go to prestigious Japanese colleges or universities after graduating, so many are academically highly motivated.

There are six English-related courses that all the students are required to take before graduating. When they are first-year students, they are required to take two of them: *English I* and *Oral Communication I*. *English I* is held three times a week, and *Oral Communication I* twice a week. The aims of *English I* are to develop students' overall basic English proficiency and to foster a positive attitude toward communication in English, in accordance with Ministry of Education, Culture, Sports, Science and Technology (MEXT) guidelines. *Oral Communication I* is aimed at developing students' basic oral and aural English abilities and fostering a positive attitude toward communicating in English. When they are second-year students, they are obliged to take *English II* and *Writing*. *English II*, which is held three times a week, focuses on developing the students' overall

English proficiency and fostering a positive attitude toward communicating in English. *Writing*, which is held once a week, is designed to develop the students' writing abilities and to foster a positive attitude toward written communication in English. When they are seniors, they need to take *Reading* and *Writing*. *Reading*, which is held three times a week, is intended to develop students' ability to understand information and the writers' intentions by reading English, and to foster a positive attitude toward communicating by utilizing these abilities. *Writing*, which meets two times a week, focuses on the same things as mentioned above for *Writing* for the second-year students.

There were six homeroom classes for a single grade level, each with approximately 40 students. Two of the six homeroom classes for second-year students were taught by the researcher four times a week in two English courses. One English course was *English II* and the other course was *Writing*. Using textbooks authorized by the MEXT in Japan is mandatory, so *New Legend English II* (Suzuki, Snowden, Eto, Kaji, Yamamoto, Ohashi, Suzuki, & Hakomori, 2003) was used for the *English II* course, and *Genius English Writing* (Sano, Matsumura, Umeda, Kawasaki, Nishizawa, & Snelling, 2004) was employed for the *Writing* course.

Sixty-nine students taught by the researcher participated in the study. Originally there were 80 students with 39 in one class (Group 1) and 41 in the other class (Group 2), but 11 students were eliminated from the study because nine of them were absent on one of the test administration days for the pretest and posttest

and two misunderstood the test instructions. Thus, Group 1 was made up of 34 students (16 male and 18 female students) and Group 2 consisted of 35 students (16 male and 19 female students). One indication of their general English proficiency was provided by the results of the Society for Testing English Proficiency (STEP) test, a well-established measure of overall English proficiency in Japan. In Group 1, there were four students at the Grade 2 level, ten students at the Grade Pre-2 level, and ten at the Grade 3 level; the proficiency level for the remaining ten was unknown. In Group 2, there were two students at the Pre-1 level, four students at the Grade 2 level, seven at the Pre-2 level, and 8 at the Grade 3 level; the proficiency level for the remaining 14 was unknown (see Table 4). STEP test results are positively correlated with the Test of English as a Foreign Language (TOEFL) (Clark & Zhang, 2005). As shown in Table 4, approximate TOEFL scores can be estimated for 14 of the participants in Group 1 and 13 of the participants in Group 2.

Table 4. *English Proficiency Levels of the Participants*

STEP Grade Level	Estimated TOEFL Score	Group 1	Group 2
Pre-1	530	0	2
2	450	4	4
Pre-2	440	10	7
3	Not specified	10	8
Unknown	Not specified	10	14

Note. The comparisons of STEP Grade Levels with estimated TOEFL scores are based on a study conducted by Clark and Zhang (2005).

With regard to informed consent, the experiment for this study was conducted in a required subject for second-year senior high school students, *English II*. Permission was received from school administrators, but in classroom studies in Japan, it is not customary to use a written informed consent form. Instead, in advance of the experiment, all the participants were orally informed of the purpose and expected duration of the study. At the same time, they were assured that all the information obtained in the process of the experiment would be kept confidential and that test scores in any measurement would not be considered in the course grading of their performance. In addition, results of the pretest and posttest would be returned to individuals with feedback on their improvement. Explanation of treatments and procedures was intentionally avoided so that participants in the different groups could not exchange information about their classroom activities.

Instrumentation

Tailored Cloze Tests

A tailored cloze test (see Appendix A) was used in order to measure the general English proficiency level of the participants and to classify them into three proficiency levels (high, intermediate, and low). The item selection for the tailored cloze test was made on the basis of item facility and item discrimination indices.

The tailored cloze tests used in this study were based on a 312-word passage and a 256-word passage taken from the reading comprehension sections of the STEP Grade Pre-2 tests, for three reasons. First, the STEP tests are widely

recognized as well-established criterion measures of overall English proficiency in Japan, and their results have a positive correlation with TOEFL (Clark & Zhang, 2005). According to Clark and Zhang's study (2005), which was conducted with 550 students from 25 countries, a STEP Grade Pre-2 level is approximately equivalent to a score of 400 on paper-based testing, and 97 on computer-based testing. Second, Grade Pre-2 levels are designed for students with an English proficiency level that Japanese senior high school juniors are expected to have achieved. This target English proficiency level for the Grade Pre-2 test applies to the participants in this study. In addition, test content for Grade Pre-2 generally is aligned with the Japanese MEXT English syllabus (Clark & Zhang, 2005). The last reason concerns the need to use topics that are almost equally familiar to the participants. The topics selected for this study were chosen in light of their familiarity to the participants and conceptual difficulty. Because the STEP tests are designed mainly for Japanese who study English, the passages are generally free from English first-language cultural bias. The topic of the 312-word passage from *The STEP TEST in Practical English Proficiency for Grade Pre-2* (The STEP Test, 2006), *Power Nap*, concerns the benefits of taking a nap. This topic is so closely related to the daily lives of the participants that the topic can be considered equally accessible to everyone. The topic of the 256-word passage from *The STEP TEST in Practical English Proficiency for Grade Pre-2* (The STEP Test, 2006), *An Animal's Best Friend*, concerns how to maintain a balance between protecting endangered animals and livestock farming in Namibia, Africa. Although the

conservation of nature is one of the most familiar topics dealt with in the study of environmental issues, the content of the passage, a Namibian conservation issue, was considered equally unknown to all of the participants.

After selection of the two passages, five types of cloze tests with word deletion patterns varying from every sixth to every tenth word were created in order to choose 40 items for the tailored cloze test. Two sentences at the beginning of the passages and more than one at the end were left intact. One hundred fifty-seven students at the same grade level were asked to complete the tests, which were randomly distributed. The word deletion patterns were counterbalanced. That is, a student who took a 312-word every-sixth-word-deletion cloze passage sat for a 256-word every-tenth-word-deletion cloze passage, and another student who took a 312-word every-tenth-word-deletion cloze passage sat for a 256-word every-sixth-word-deletion cloze passage. All scoring was done by the exact word scoring procedure, in which only the original word in the passage is counted as correct. All the items used for both cloze tests were statistically analyzed using item facility and item discrimination. Item facility is a statistical index that is used for examining the percentage of students who answer a given item correctly. Item discrimination is another statistical index that is used to indicate the degree to which an item separates the students who performed well from those who performed poorly (Brown, 1996). Items were chosen so that they were no closer than five words and no further than fifteen words from the nearest item. Statistically, ideal items are those which have an average item facility of .50 and the highest available item

discrimination, but there is little chance that items have an item facility of exactly .50, so items within a item facility range between .30 and .70 are usually considered acceptable (Brown, 1996). After those items within the allowable range of item facilities were identified, the items among them showing the highest item discrimination indexes were selected for retention.

The results of the item analysis showed that the 20 items selected from the 312-word passage cloze patterns have an item facility of .32 to .73 and an item discrimination of .36 to .82 and the 20 items selected from the 256-word passage cloze patterns had an item facility of 0.17 to .70 and item discrimination of .33 to 1.00. One problem in the process of item selection is that the two items selected from the 256-word passage cloze tests had an item facility of 0.17 and 0.27; however, they were used because they had the best item facility and a high item discrimination among the remaining items. The Cronbach alpha internal consistency reliability estimate was .85 for the two passages combined. The 40-item tailored cloze tests were administered to the participants at the beginning of the five-month study as one of the indicators of general proficiency used to divide the participants into three proficiency groups.

Short-answer Tests

The participants' reading comprehension ability was estimated using a short-answer question, limited production response type (Bachman & Palmer, 1996). In this testing technique, the participants are asked a question that demands an answer

varying from one word to one or two sentences. The justification for this technique is that the test-takers are required to produce an answer to the short-answer questions instead of simply choosing a correct option or filling in blanks; this enables the researcher to inspect the written responses to see whether the test-takers understood the content of the passage (Heaton, 1988). In addition, short-answer questions force students to interact with a text while they are thinking of the responses (Carrell, 1988; Kembo, 2001; Nuttall, 2005). For these reasons, short-answer questions have been considered to be the best way of testing in-depth information processing, such as the integration of different propositions and appropriate world knowledge (Kembo, 2001).

In order to remove ambiguities in the wording of the questions and to write questions that would elicit specific answers, the questions were piloted with 77 students in the same senior high school: one passage was a 321-word passage from *The STEP in Practical English Proficiency for Grade Pre-2* (The STEP Test, 2006) and the other was a 341-word passage from *From Reader to Reading Teacher* (Aebersold & Field, 1997), a research and practice book on effective approaches to language teaching.

The selection of the two passages was made using the same criteria as those used to select the *Tailored Cloze Tests*. That is, the two passages were selected because of the participants' familiarity with the topics and the perception that they were at an appropriate level of difficulty in terms of lexis, grammar, and text length. The pilot tests were made up of a set of 18 reading questions for each of the two

passages. The number of reading questions was reduced to 15 for each passage in order to make the number of textually explicit (TE) and inferential (IF) questions equal. In this way, the 30-item reading comprehension tests (15 TE and 15 IF questions) based on the 321-word and 341-word passages were created and administered to the participants at the beginning and end of the five-month classroom research.

Model answers were determined in the following process. First, two English speaking informants who are from the United States and had worked as English teachers in Japanese senior high schools for more than ten years took the tests. They discussed their own answers and the answers obtained from the pilot testing with the researcher until reaching an agreement on acceptable answers. The results of the two discussions on acceptable answers determined model answers for all of the questions.

Interrater reliability was determined with the assistance of one of the English-speaking teachers who participated in decision-making on model answers. First, the teacher and the researcher individually marked two sample answer sheets from the pretest for practice. Then, when the marking was finished, the marks were compared. When discrepancies were found, the reasons for the differences were discussed and agreement was reached, while referring to the guidelines. Finally, following the preliminary process of achieving scoring consistency, the markers went through fifteen answer sheets that were randomly selected from the pretest. Interrater reliability, using Cohen's kappa was calculated and the obtained kappa

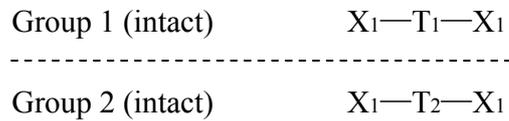
value was .76. Because this value is greater than the widely accepted value of .70, it was concluded that interrater reliability was satisfactory.

Marking guidelines were as follows. Responses with all required information in acceptable language were awarded 2 points, responses with the core but not sufficient information or sufficient information but not written in acceptable language were awarded 1 point, and responses with no expected information or that were grammatically unacceptable were awarded no points.

Research Design

An alternative treatment pretest and posttest with nonequivalent intact groups design was used in this study. This classroom research involved the use of two classes where participants had already been assigned on the basis of their junior high school academic grades and scores on the entrance exam to get into senior high school. For this reason, in this research it was impossible to select students randomly to begin with. This research design included a pretest measure: the short-answer tests with 30 question items, which are indicated as X_1 in the schematic representation for the research described below. After the administration of the pretest came different treatments (instruction with an emphasis on IF reading questions for Group 1 and instruction with an emphasis on TE reading questions for Group 2), which are indicated as T_1 and T_2 in the schematic representation described below. At the end of the experiment of five months, a posttest measure (the same short-answer tests as conducted for the pretest measure) was

administered for both groups. The following diagram illustrates the schematic representation for the research design.



Question Types

Textually explicit reading questions are text-bound questions whose answers are locally and explicitly stated in the text. They require readers to read in order to locate and comprehend discrete pieces of information such as facts, names, dates, times, locations, lexical items, and propositions found in a single sentence or adjacent sentences. They can be usually answered using words from the text due to a lexical overlap with both question wording and correct answer (Cohen & Upton, 2006; Enright, et al., 2000; Nuttall, 2005; Raphael, 1982). Therefore, they are labeled Textually Explicit Questions (Pearson & Johnson, 1978) or Right There (Raphael, 1982, 1986). Textually explicit reading questions involve not only recognition and recall but also reorganization and reinterpretation of text information, which is similar to Type 2 in Nuttall's question category (2005). That is, readers are required to reinterpret the literal information found in the text or to obtain bits and pieces of surface information from adjacent sources of the text and put them together. A numerical calculation involved in the process of arriving at answers is one example of reinterpretation of text information (see question 2 for Passage 1 and question 10 for Passage 2 in the short-answer tests in Appendix B).

Locating relevant pieces of information from two adjacent sentences and putting them together on the way to answering are an example of reorganization of text information (see question 15 for Passage 1 in the short-answer tests in Appendix B).

In contrast, inferential questions are more knowledge-bound or reader-bound questions whose answers largely depend on readers' cognitive resources such as relevant linguistic knowledge, background knowledge, world knowledge, or even context in the process of arriving at answers. They require readers to search more globally for text information in various parts of the text and to infer factual information, main ideas, comparisons, and cause-effect relationships unstated by the writer.

Inferential questions whose answers are found on the page but are not explicitly expressed are labeled Textually Implicit Questions (Pearson & Johnson, 1978) or Search-and-Think (Raphael, 1982, 1986). For example, inferential questions involve identification of the referents of pronominals, pro-forms or ellipsis like questions 8, 9, and 11 for Passage 1 and questions 3 and 12 for Passage 2 in the short-answer test in Appendix B. They also include establishment of local coherence between the contents of adjacent text constituents like a phrase, proposition, and clause or between short sequences of constituents (see question 14 for Passage 1 and question 11 for Passage 2 in Appendix B).

Inferential questions whose answers largely reside in the readers' mind are labeled Scriptally Implicit Questions (Pearson & Johnson, 1978) or On My Own (Raphael, 1982, 1986). They involve relating the reading passage to real life such

as readers' own opinions, knowledge, imagination and experience (see question 7 for Passage 1 and questions 1 and 15 for Passage 2 in Appendix B).

Procedures

The study was conducted over five months with the pretest administration at the beginning (week one) and the posttest administration at the end (week seventeen). In week one, at the beginning of the different treatments, the participants were allowed 50 minutes to take the tailored cloze tests (see Appendix A) and the short-answer reading comprehension tests (see Appendix B). According to the participants' scores on the tailored cloze tests and their English course grades, they were classified into three proficiency levels: high-proficiency, intermediate-proficiency, and low-proficiency. Participants ranked within the top one-third on the continuum of the combined scores of the cloze tests and the course grades were classified as high proficiency. Participants within the middle one-third were classified as intermediate proficiency. Participants with scores in the lowest one-third were classified as low proficiency. The descriptive statistics for the cloze tests and course grades for Groups 1 and 2 are shown in Table 5. Between week one and week 17, Groups 1 and 2 received different treatments in the same English course, *English II* (see its course syllabus in Appendix C). Group 1 was given a worksheet that had a set of reading questions with three or four IF questions and one TE question plus a summary writing task (Appendix D). In contrast, Group 2 was provided with a worksheet that had another set of reading questions with one IF

question and three or four TE questions plus the same summary writing task given for Group 1 (Appendix D). In this way the two groups were engaged in the reading tasks with the different sets of the two types of reading questions, TE and IF questions, when they read a new part or section of one article in the textbook *New Legend English II* (Suzuki, et al., 2003). All classroom activities except for the different sets of reading questions and the related interaction between the researcher and participants to confirm answers to the assigned questions were intended to be carried out in the same manner. The worksheets were always collected at the end of each lesson in order to monitor the students' learning and the extent to which the students had understood the reading material. The collected worksheets were regularly returned to the participants in the following lesson with an assessment of their original answers to the assigned reading questions and their completed summary writing accompanied by brief comments on their performance.

In week seventeen, the posttest was administered to both groups using the same procedures as used on the pretest. Both groups were allowed 50 minutes to complete two cloze tests (Appendix A) and two short-answer reading comprehension tests (Appendix B). The passages and questions were exactly the same except that the passages for both the cloze and the reading comprehension tests were counterbalanced with those in the pretest in order to avoid a practice effect.

Table 5. *Descriptive Statistics for Cloze Tests and Course Grades for Each Group of Participants*

		Group 1 <i>n</i> = 34	Group 2 <i>n</i> = 35
Cloze Tests	<i>M</i>	17.53	14.29
	<i>SE</i>	1.08	1.43
	95% CI Lower Bound	15.34	11.39
	95% CI Upper Bound	19.72	17.18
	<i>SD</i>	6.27	8.43
	Skewness	.28	.36
	<i>SES</i>	.40	.40
	Kurtosis	-1.00	-.64
	<i>SEK</i>	.79	.78
	Course Grades	<i>M</i>	28.72
<i>SE</i>		.81	.95
95% CI Lower Bound		27.07	25.93
95% CI Upper Bound		30.37	29.80
<i>SD</i>		4.72	5.65
Skewness		.01	-.21
<i>SES</i>		.40	.40
Kurtosis		-1.30	-.65
<i>SEK</i>		.79	.78

CHAPTER 4

RESULTS

Data Analysis

In order to investigate the research questions, three three-way repeated-measures ANOVAs were performed with three independent variables on each of three dependent variables. The three independent variables consist of treatment group and proficiency as between-subjects variables, and test administration time as a within-subjects variable. Treatment group for the between-subjects variable denotes two treatment groups. One of them is the group (Group 1) that received a set of three or more IF plus one TE reading questions. The other is the group (Group 2) that received another set of three or more TE plus one IF reading questions. Proficiency for the other between-subjects variable denotes three proficiency levels: high, intermediate and low. Test administration time, the within-subjects variable, indicates two tests: the pretest and posttest. In order to make the statistical analyses easier, test administration time and treatment group effects were analyzed before proficiency effects for the three independent variables. The three dependent variables are the total scores of (a) both TE and IF reading questions, (b) only TE reading questions, and (c) only IF reading questions. The main and interaction effects for performance on reading questions were examined in the order mentioned above.

Before investigating the main factor and interaction effects, the descriptive statistics for the participants' performance on the pretest and posttest were produced and shown in tables. Levene's test was conducted to check the assumption of homogeneity of variances for the two levels of the repeated-measures variables: in this case the pretest and the posttest. When multiple ANOVAs are conducted, it is necessary to make a Bonferroni adjustment in order to avoid a Type I error approximately. Since three three-way repeated-measures ANOVAs were done in this study, a Bonferroni adjustment was made by dividing the alpha level of .05 by three (the number of comparisons) and an adjusted alpha level of .017 was employed.

Test Administration Time and Treatment Effects on Reading Questions
*Test Administration Time and Treatment Effects on Textually Explicit and
Inferential Reading Questions*

In order to examine the main effects of test administration time and treatment group on the total scores of TE and IF reading questions, the descriptive statistics for performance on the pretest and posttest for two treatment groups were calculated (see Table 6).

The main effects of test administration time and treatment group on the total scores of both TE and IF reading questions in the pretest and posttest were statistically examined with the three-way repeated-measures ANOVA. The independent variables were the within-subjects variable of test administration time

(pretest and posttest) and the between-subjects variables of treatment group (Groups 1 and 2) and proficiency (high, intermediate and low). The assumption of homogeneity of variances was checked using Levene's test. Table 7 showing the

Table 6. *Descriptive Statistics for Textually Explicit and Inferential Reading Questions on the Pretest and Posttest for Groups 1 and 2*

		Group 1 <i>n</i> = 34	Group 2 <i>n</i> = 35
Pretest	<i>M</i>	32.68	35.29
	<i>SE</i>	1.72	1.54
	95% CI Lower Bound	29.18	32.17
	95% CI Upper Bound	36.18	38.41
	<i>SD</i>	10.03	9.08
	Skewness	-.61	-.34
	<i>SES</i>	.40	.40
	Kurtosis	-.25	-.14
	<i>SEK</i>	.79	.78
	Posttest	<i>M</i>	40.06
<i>SE</i>		.98	1.39
95% CI Lower Bound		38.06	35.95
95% CI Upper Bound		42.06	41.60
<i>SD</i>		5.74	8.22
Skewness		-.65	-1.18
<i>SES</i>		.40	.40
Kurtosis		.74	2.36
<i>SEK</i>		.79	.78

Table 7. *Levene's Test of Equality of Error Variances*

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
pretest	.879	5	63	.500
posttest	1.010	5	63	.420

Note. Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Levene's test result indicates that the assumption of homogeneity of variances was met ($p > .05$). The ANOVA summary, which is shown in Table 8, revealed a statistically significant difference only for test administration time, $F(1, 63) = 19.709, p < .017$, thus indicating that the participants performed significantly better on the posttest than on the pretest. The strength of association was estimated using eta squared (see Table 8). The main effect for test administration time accounted for 24% of the variance in the total scores of both TE and IF reading questions. There was not a significant main effect for treatment group. No significant interaction effect was found between test administration time and treatment group, either. However, Power was quite low: .200 for treatment group and .316 for the interaction. Therefore, these two results must be interpreted with caution.

With regard to the skewness, the distribution of test scores in the posttest for Group 2 was negatively skewed. According to Brown (1996, 1997), values of two times larger than the standard error of skewness (SES) are probably skewed to a significant degree. The standard error of the skewness for Group 2's posttest was .40 (see Table 6). Since two times the standard error of the skewness is .80 and the absolute value of the skewness statistic for Group 2's posttest is 1.18, which is greater than .80, it is assumed that the distribution was significantly skewed. In other words, the mass of the distribution of posttest scores for Group 2 was concentrated toward the top of the scale (60 points).

The mean scores gained for Groups 1 and 2 for the pretest and posttest, which are shown in Table 6 for the descriptive statistics, are plotted in Figure 1. As Figure

1 shows, although there was no significant interaction between test administration time and treatment group, the descriptive statistics seem to suggest that Group 2 had a higher mean than Group 1 on the pretest, while Group 1 performed better and had a higher mean than Group 2 on the posttest as indicated by Group 1's larger gain scores (7.38) than Group 2's gain scores (3.48).

Table 8. *Repeated-Measures ANOVA for Textually Explicit and Inferential Reading Questions*

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	Power	<i>eta</i> ²
Between subjects							
Treatment Group (G)	87.051	1	87.051	1.281	.262	.200	.020
Proficiency (P)	1469.008	2	734.504	10.806	.000	.987	.255
G x P	511.570	2	255.785	3.763	.029	.667	.107
Error	4282.100	63	67.970				
Within subjects							
Test Administration Time (TAT)	985.743	1	985.743	19.709	.000	.992	.238
TAT x G	113.061	1	113.061	2.261	.138	.316	.035
TAT x P	21.177	2	10.589	.212	.810	.082	.007
TAT x G x P	5.721	2	2.860	.057	.944	.058	.002
Error	3150.868	63	50.014				

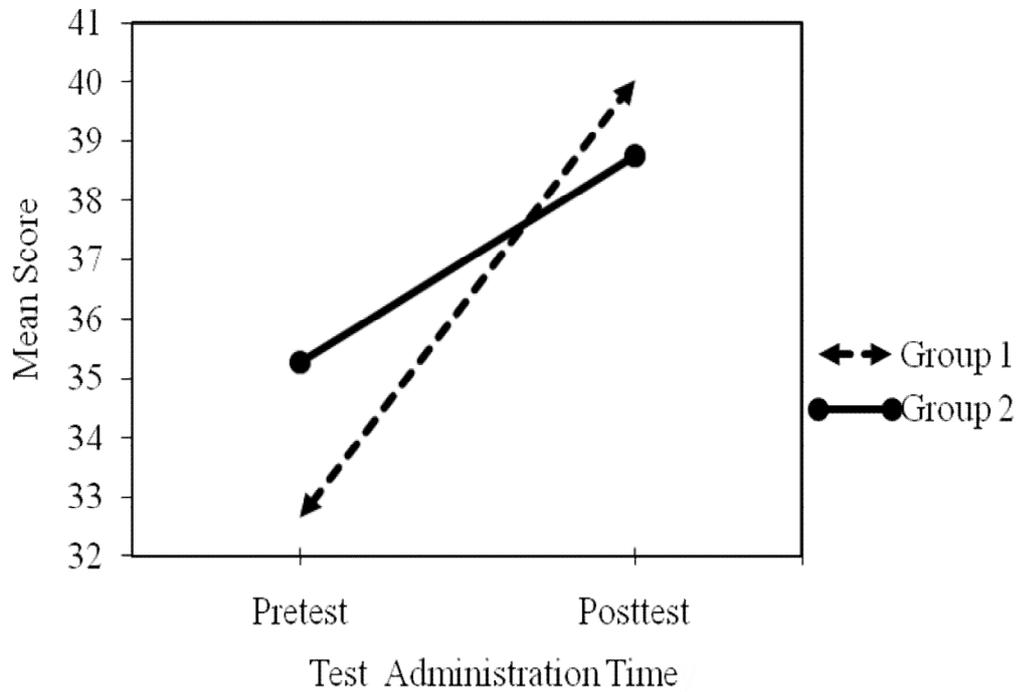


Figure 1. The relationship between test administration time and treatment group for textually explicit and inferential questions.

Test Administration Time and Treatment Effects on Textually Explicit Reading Questions

Following the same procedures used in the TE and IF reading questions analysis, the effects of test administration time and treatment group on the participants' performance on only TE reading questions were examined. Table 9 displays the descriptive statistics for the performance of Groups 1 and 2 on TE reading questions.

Table 9. *Descriptive Statistics for Textually Explicit Reading Questions on the Pretest and Posttest for Groups 1 and 2*

		Group 1 <i>n</i> = 34	Group 2 <i>n</i> = 35
Pretest	<i>M</i>	21.68	22.57
	<i>SE</i>	1.07	.80
	95% CI Lower Bound	19.51	20.94
	95% CI Upper Bound	23.84	24.20
	<i>SD</i>	6.21	4.75
	Skewness	-1.00	-.95
	<i>SES</i>	.40	.40
	Kurtosis	-.29	.66
	<i>SEK</i>	.79	.78
	Posttest	<i>M</i>	25.47
<i>SE</i>		.48	.75
95% CI Lower Bound		24.49	23.14
95% CI Upper Bound		26.45	26.18
<i>SD</i>		2.82	4.42
Skewness		-1.45	-2.26
<i>SES</i>		.40	.40
Kurtosis		3.21	6.44
<i>SEK</i>		.79	.78

A three-way repeated-measures ANOVA was carried out with three independent variables and one dependent variable. The independent variables were test administration time, treatment group and proficiency. The dependent variable was the total scores of TE reading questions. Levene's test was conducted to check the assumption of homogeneity of variances for two levels of the repeated-measures variable (pretest and posttest), and the assumption was met ($p > .05$), as displayed in Table 10.

Table 10. *Levene's Test of Equality of Error Variances*

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
pretest	1.971	5	63	0.95
posttest	2.180	5	63	0.68

Note. Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

The ANOVA summary, which is shown in Table 11, revealed a statistically significant difference for test administration time, $F(1, 63) = 12.95, p < .017$, thus indicating that the participants performed in different ways on the two tests. The strength of association was estimated using eta squared (see Table 11). The main effect of test administration time explained 17% of the variance in TE reading questions. There was no significant main effect of treatment group, nor was there any significant interaction effect between test administration time and treatment group. However, Power was very low: .074 for treatment group and .172 for the interaction between test administration time and treatment group. Therefore, caution must be taken when interpreting these two results.

In terms of the skewness, the distribution of test scores on the pretest and posttest for Groups 1 and 2 were negatively skewed. Brown (1996, 1997) says that values of two times larger than the standard error of skewness (SES) are probably skewed to a significant degree. The standard error of the skewness for the pretest and posttest for both Groups 1 and 2 was .40 (see Table 9). Two times the standard error of the skewness is .80. The absolute values of the skewness statistic were: 1.00 for Group 1's pretest, .95 for Group 2's pretest, 1.45 for Group 1's posttest,

and 2.26 for Group 2's posttest. All these absolute values are greater than .80. Therefore, it is assumed that the distribution of test scores on the pretest and posttest for both Groups 1 and 2 were significantly skewed. In other words, the mass of the distribution of test scores in the pretest and posttest scores for both Groups 1 and 2 were concentrated toward the top of the scale (30 points).

The mean scores on the pre-and posttest for Groups 1 and 2 in Table 9 are plotted in Figure 2. As Figure 2 shows, although there was no significant interaction between test administration time and treatment group, the descriptive statistics seem to suggest that Group 2 slightly outperformed Group 1 on the pretest, while Group 1 performed better than Group 2 to a slight degree on the posttest due to Group 1's larger gain scores (3.79) than Group 2's (2.09).

Table 11. *Repeated-Measures ANOVA for Textually Explicit Reading Questions*

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	Power	<i>eta</i> ²
Between subjects							
Treatment Group (G)	4.394	1	4.394	.212	.646	.074	.003
Proficiency (P)	161.158	2	80.579	3.896	.025	.683	.110
G x P	40.407	2	20.203	.977	.382	.213	.030
Error	1302.99	63	20.682				
Within subjects							
Test Administration Time (TAT)	295.642	1	295.642	12.950	.001	.943	.171
TAT x G	23.991	1	23.991	1.051	.309	.172	.016
TAT x P	.299	2	.149	.007	.993	.051	.000
TAT x G x P	12.040	2	6.020	.264	.769	.090	.008
Error	1438.303	63	22.830				

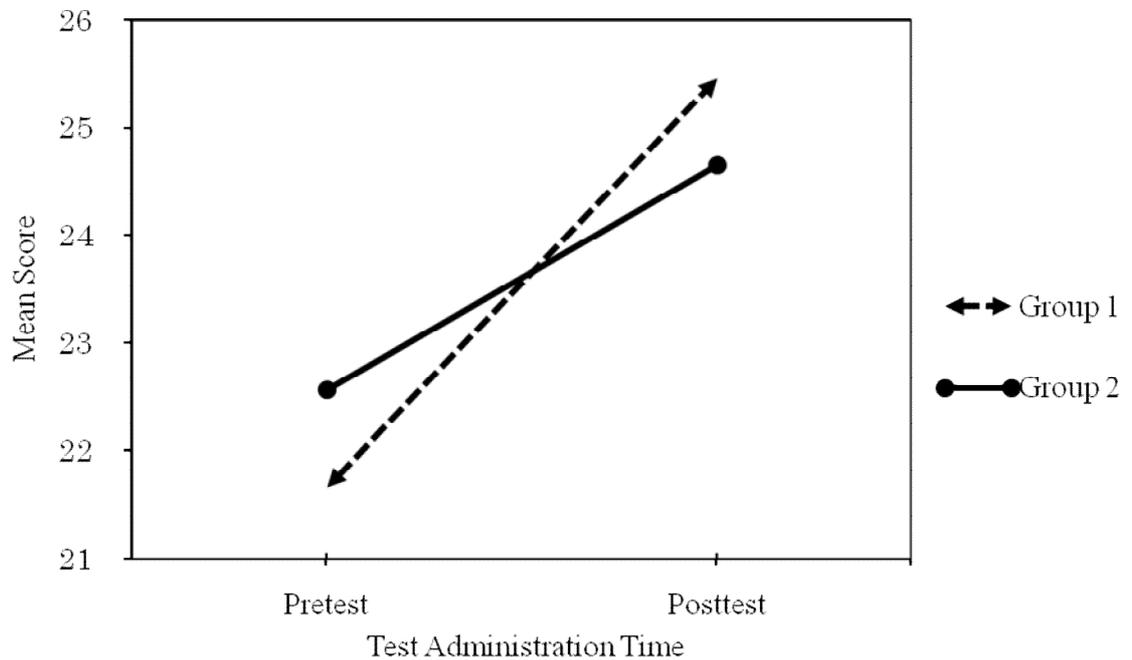


Figure 2. The relationship between test administration time and treatment group for textually explicit reading questions.

Test Administration Time and Treatment Effects on Inferential Reading Questions

Following the same procedures adopted in the previous two sections, the effects of test administration time and treatment group on performance on only IF reading questions were investigated. Table 12 shows the descriptive statistics for performance on the IF reading questions for Groups 1 and 2.

A three-way repeated-measures ANOVA was carried out with test administration time, treatment group and proficiency as the independent variables and the total scores of IF reading questions as the dependent variable. Levene's test

Table 12. *Descriptive Statistics for Inferential Reading Questions on the Pretest and Posttest for Groups 1 and 2*

		Group 1 <i>n</i> = 34	Group 2 <i>n</i> = 35
Pretest	<i>M</i>	11.00	12.71
	<i>SE</i>	.80	.92
	95% CI Lower Bound	9.38	10.84
	95% CI Upper Bound	12.62	14.59
	<i>SD</i>	4.64	5.45
	Skewness	.31	.38
	<i>SES</i>	.40	.40
	Kurtosis	.15	-.29
	<i>SEK</i>	.79	.78
	Posttest	<i>M</i>	14.59
<i>SE</i>		.63	.86
95% CI Lower Bound		13.31	12.37
95% CI Upper Bound		15.87	15.86
<i>SD</i>		3.68	5.07
Skewness		-.01	-.06
<i>SES</i>		.40	.40
Kurtosis		-1.14	-1.10
<i>SEK</i>		.79	.78

Table 13. *Levene's Test of Equality of Error Variances*

	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
pretest	1.971	5	63	.514
posttest	2.180	5	63	.237

Note. Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

was used to test the assumption of homogeneity of variance for two levels of the repeated-measures variable (pretest and posttest). The result indicates that

variances were homogeneous for both levels of repeated-measures variables because the significance values were greater than .05 (see Table 13).

The ANOVA summary shown in Table 14 revealed a statistically significant difference for test administration time, $F(1, 63) = 19.504, p < .017$. This result indicates that the participants performed significantly better on the posttest than on the pretest. The strength of association was estimated using eta squared. The main effect for the test administration time accounted for 24% of the variance in IF reading questions. There was not a significant main effect of group, nor was there a significant interaction effect between test administration time and treatment group. However, as Power was quite low, .324 for treatment group and .419 for the interaction, these two results must also be interpreted with caution.

The means gained for Groups 1 and 2 on the pretest and posttest are plotted in Figure 3. As Figure 3 shows, although there was no significant interaction between test administration time and treatment group, the descriptive statistics seem to suggest that Group 2 performed slightly better than Group 1 on the pretest, whereas Group 1 outperformed Group 2 to a slight degree on the posttest because Group 1 gained 3.59 points and Group 2 gained 1.40 points.

Table 14. *Repeated-Measures ANOVA for Inferential Reading Questions*

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	Power	<i>eta</i> ²
Between subjects							
Treatment Group (G)	52.329	1	52.329	2.330	.132	.324	0.36
Proficiency (P)	662.656	2	331.328	14.756	.000	.999	.319
G x P	264.465	2	132.233	5.8989	.005	.860	.158
Error	1414.629	63	22.454				
Within subjects							
Test Administration Time (TAT)	201.705	1	201.705	19.504	.000	.992	.236
TAT x G	32.890	1	32.890	3.180	.079	.419	.048
TAT x P	17.904	2	8.952	.866	.426	.192	.027
TAT x G x P	4.830	2	2.415	.234	.729	.085	.007
Error	651.517	63	10.342				

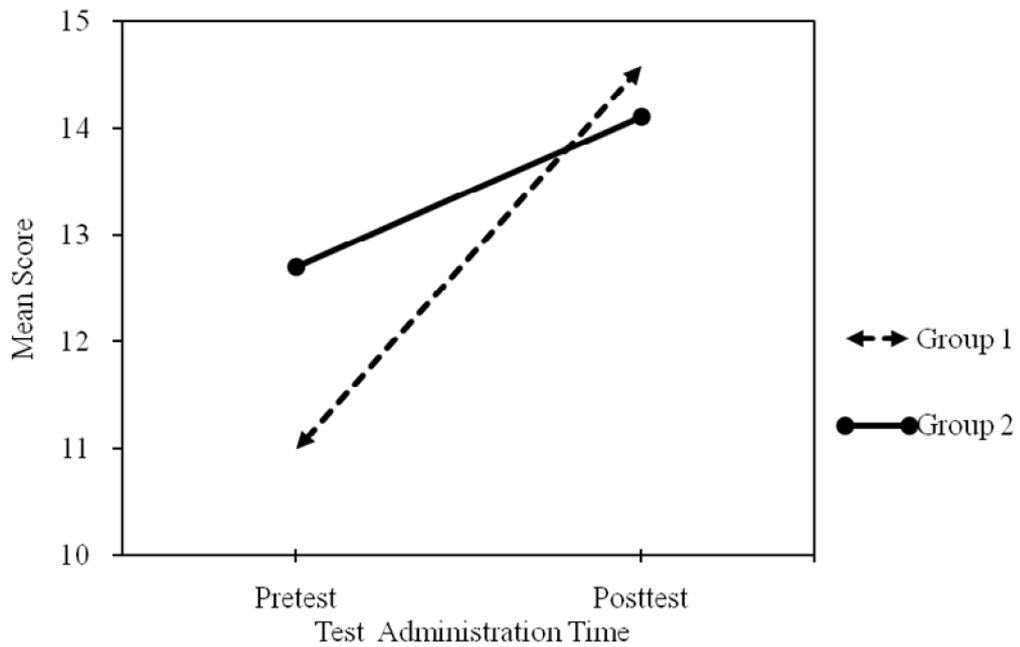


Figure 3. The relationship between test administration time and treatment group for inferential reading questions.

Proficiency Effects on Reading Questions

In order to examine the main effect of proficiency (high, intermediate, and low) on performance on the short-answer tests, the descriptive statistics for performance on the reading questions for three proficiency levels in each group were produced and shown in tables. The assumption of homogeneity of variances was already checked and met. The main effect of proficiency is displayed in the related ANOVA summaries in the previous section entitled *Test Administration Time and Treatment Effects on Reading Questions*. The proficiency effect on the total scores of (a) both TE and IF reading questions, (b) only TE reading questions, and (c) only IF reading questions will be reported in this order.

Proficiency Effects on Textually Explicit and Inferential Reading Questions

First, the participants' performance on both TE and IF reading questions for three proficiency levels will be reported. Table 15 shows the descriptive statistics for performance on TE and IF reading questions on the pretest and posttest for three proficiency levels in Groups 1 and 2. As two participants that were assigned to the high proficiency group based on combined scores of the cloze test and the English course grade performed poorly on the pretest, and there was also a small sample size in each proficiency category, the means for the high and intermediate proficiency participants in Group 1 were almost the same.

The ANOVA summary displayed in Table 8 revealed the significant main effect of proficiency, $F(2,63) = 10.806, p < .017$. The strength of association was estimated using eta squared. The main effect of proficiency explained 26% of the variance in the total scores of TE and IF reading questions. However, there was not a significant interaction between test administration time and proficiency or between treatment group and proficiency. These results show that proficiency significantly affected how participants performed in TE and IF reading questions.

In order to identify the precise location of the proficiency differences, estimated marginal means for this main effect of proficiency were calculated. Table 16 shows the estimated marginal means for the main effect of proficiency. Post hoc multiple comparisons of marginal means using the Bonferroni test were performed. Table 17 displays the results of the multiple comparisons for the main proficiency effect. The results revealed that there were statistically significant differences

Table 15. *Descriptive Statistics for Textually Explicit and Inferential Reading*

Questions on the Pretest and Posttest for Three Proficiency Levels in Groups 1 and 2

		Group 1			Group 2		
		Low <i>n</i> = 9	Inter. <i>n</i> = 12	High <i>n</i> = 13	Low <i>n</i> = 14	Inter. <i>n</i> = 11	High <i>n</i> = 10
Pretest	<i>M</i>	31.00	33.33	33.23	30.29	35.09	42.50
	<i>SE</i>	2.89	3.13	2.98	2.36	2.11	2.19
	95% CI						
	Lower Bound	24.33	26.44	26.73	25.18	30.39	37.55
	95% CI						
	Upper Bound	37.67	40.23	39.73	35.39	39.79	47.45
	<i>SD</i>	8.68	10.85	10.76	8.84	6.99	6.92
	Skewness	-.55	-1.66	.05	-.23	-.61	-.18
	<i>SES</i>	.72	.64	.62	.60	.66	.69
	Kurtosis	-1.43	2.31	-1.44	-.33	-.48	-.60
	<i>SEK</i>	1.19	1.23	1.19	1.15	1.28	1.33
Posttest	<i>M</i>	37.00	40.58	41.69	32.86	39.55	46.20
	<i>SE</i>	1.16	2.06	1.40	2.15	1.65	1.11
	95% CI						
	Lower Bound	34.34	36.05	38.64	28.22	35.86	43.68
	95% CI						
	Upper Bound	39.66	45.12	44.75	37.49	43.23	48.72
	<i>SD</i>	3.46	7.14	5.06	8.03	5.48	3.52
	Skewness	1.18	-.95	-1.72	-1.60	-.86	-.44
	<i>SES</i>	.72	.64	.62	.60	.66	.69
	Kurtosis	1.03	1.59	3.95	3.11	.83	-1.10
	<i>SEK</i>	1.40	1.23	1.19	1.15	1.28	1.33

between high proficiency and low proficiency participants, and between

intermediate proficiency and low proficiency participants.

The estimated means for the three proficiency levels in Groups 1 and 2 are plotted in Figure 4. Figure 4 shows that Group 2 scored higher in participants with high proficiency than Group 1, Groups 1 and 2 scored almost the same for

Table 16. *The Estimated Marginal Means for the Main Effect of Proficiency*

Group	Proficiency	Mean	SE	95% Confidence Interval	
				Lower Bound	Upper Bound
G1	High	37.46	1.62	34.23	40.69
	Intermediate	36.96	1.68	33.60	40.32
	Low	34.00	1.94	30.12	37.88
G2	High	44.35	1.84	40.67	48.03
	Intermediate	37.32	1.76	33.81	40.83
	Low	31.57	1.56	28.46	34.69

Note. G1 = Group 1; G2 = Group 2.

Table 17. *Bonferroni Test of Multiple Comparisons of Proficiency Levels*

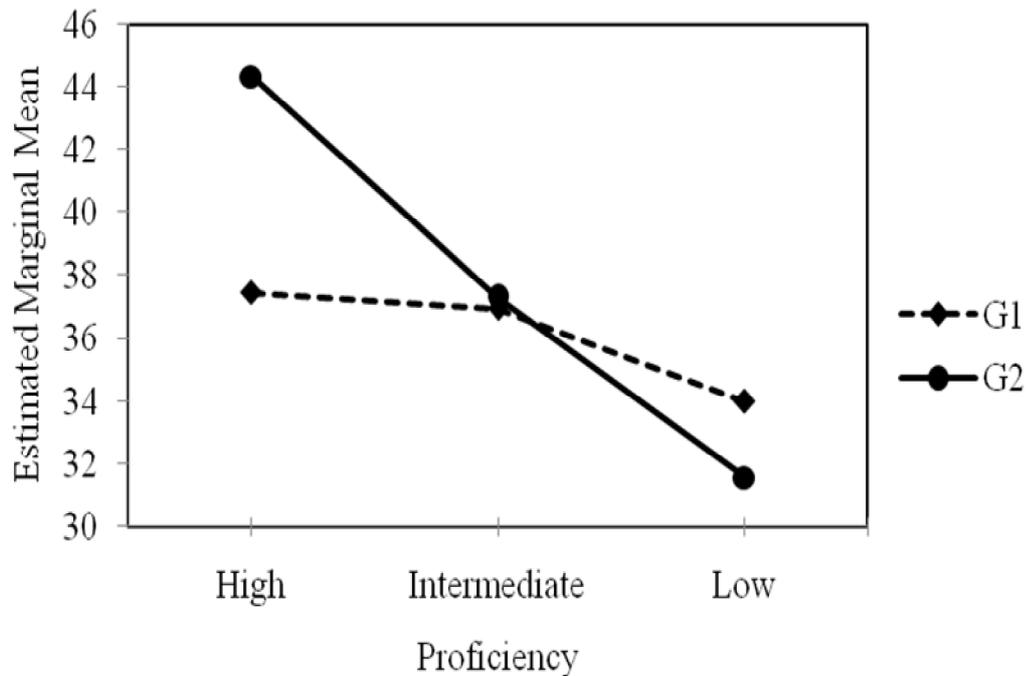
(I) prof	(J) prof	Mean Difference (I - J)	SE	p	95% Confidence Interval for Difference	
					Lower-Bound	Upper-Bound
1	3	8.12*	1.75	.000	3.82	12.42
2	3	4.35*	1.74	.045	.07	8.64

Note. Based on estimated marginal means. *The mean difference is significant at

the .05 level. Adjustment for multiple comparisons: Bonferroni. 1 = high proficiency level; 2 = intermediate proficiency level; 3 = low proficiency level.

participants with intermediate proficiency, and Group 1 performed slightly better in participants with low proficiency than Group 2. However, the ANOVA result did not indicate a statistically significant interaction effect for proficiency by treatment group. In addition, although Groups 1 and 2's performance decreased as the proficiency level fell, this fall was more pronounced for Group 2. As a result, the main effect of proficiency seems to reflect that proficiency mainly influenced how Group 2 performed on TE and IF reading questions and that the higher

proficiency participants performed significantly better than the low proficiency participants.



Note. G1 = Group 1; G2 = Group 2.

Figure 4. The relationship between group and proficiency for textually explicit and inferential reading questions.

Proficiency Effects on Textually Explicit Reading Questions

Next, the participants' performance on TE reading questions will be reported. The descriptive statistics for performance on TE reading questions on the pretest and posttest for the three proficiency levels in Groups 1 and 2 are shown in Table 18.

Table 18. *Descriptive Statistics for Textually Explicit Reading Questions on the Pretest and Posttest for Three Proficiency Levels in Groups 1 and 2*

		Group 1			Group 2		
		Low <i>n</i> = 9	Inter. <i>n</i> = 12	High <i>n</i> = 13	Low <i>n</i> = 14	Inter. <i>n</i> = 11	High <i>n</i> = 10
Pretest	<i>M</i>	20.56	22.42	21.77	21.00	22.45	24.90
	<i>SE</i>	1.92	1.99	1.72	1.58	1.14	.94
	95% CI						
	Lower Bound	16.12	18.03	18.02	17.58	19.92	22.78
	95% CI						
	Upper Bound	24.99	26.81	25.52	24.42	24.99	27.02
	<i>SD</i>	5.77	6.91	6.21	5.92	3.78	2.96
	Skewness	-.78	-1.68	-.75	-.58	-.45	-.75
	<i>SES</i>	.72	.64	.62	.60	.66	.69
	Kurtosis	-1.59	2.03	-1.11	-.32	-1.34	.31
	<i>SEK</i>	1.40	1.23	1.19	1.15	1.28	1.33
Posttest	<i>M</i>	24.67	25.33	26.15	22.64	25.36	26.70
	<i>SE</i>	.71	1.00	.72	1.58	.89	.50
	95% CI						
	Lower Bound	23.04	23.13	24.58	19.24	23.39	25.58
	95% CI						
	Upper Bound	26.30	27.54	27.73	26.04	27.34	27.82
	<i>SD</i>	2.12	3.47	2.61	5.89	2.94	1.57
	Skewness	-.85	-1.71	-1.83	-1.64	-.24	-1.55
	<i>SES</i>	.72	.64	.62	.60	.66	.69
	Kurtosis	-.85	4.92	4.28	2.52	-1.57	2.89
	<i>SEK</i>	1.40	1.23	1.19	1.15	1.28	1.33

Note. Inter = Intermediate.

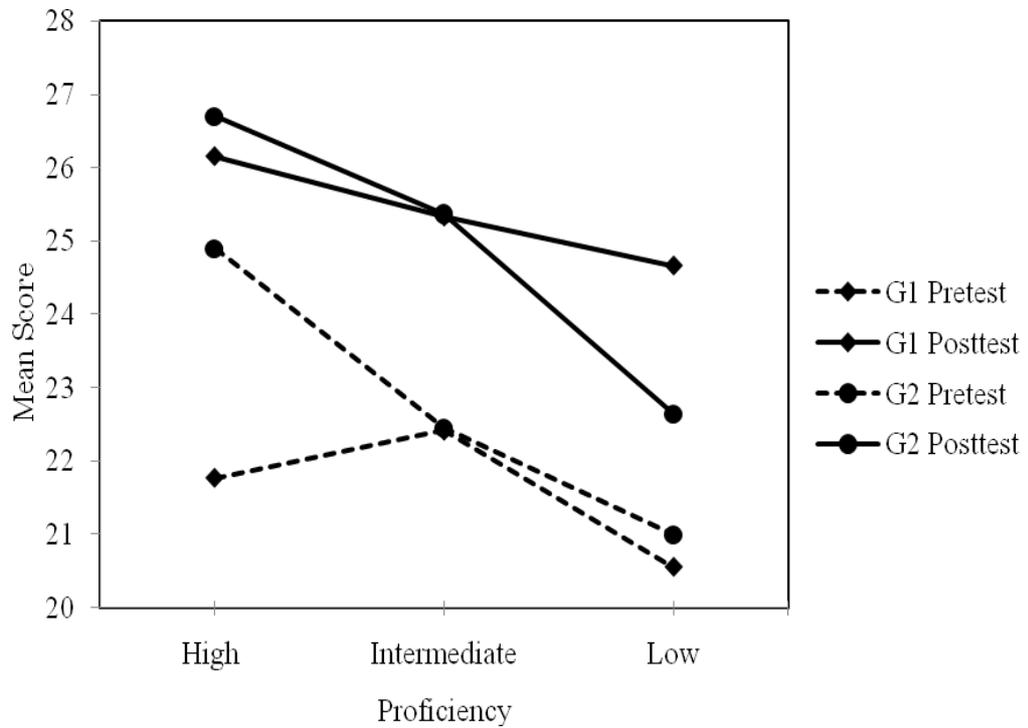
The ANOVA summary, which is shown in Table 11, indicates that there was not a significant main effect of proficiency. There were not any significant interaction effects found between test administration time and proficiency or between treatment group and proficiency. Power for proficiency was .683, which was rather low. Power was quite low for the interactions: .051 for test administration time by proficiency, and .213 for treatment group by proficiency.

Therefore, these results must be interpreted with caution.

The mean scores gained for the three proficiency levels in Groups 1 and 2 shown in Table 18 are plotted in Figure 5. Figure 5 shows that only the high proficiency participants in Group 2 scored higher than their counterparts in Group 1, and that participants with intermediate and low proficiency in the two groups scored more or less the same on the pretest.

The participants with all proficiency levels in the two groups performed better on the posttest than on the pretest. On the posttest participants with high and intermediate proficiency in the two groups had almost the same means, and only the low proficiency participants in Group 1 scored higher means than their counterparts in Group 2.

In terms of mean gain scores over the two tests, the high proficiency participants in Group 1 showed the greatest gain (4.38), while their counterparts in Group 2 showed the second smallest gain (1.80). Participants with intermediate proficiency in both Groups 1 and 2 had exactly the same gain (2.91). The low proficiency participants in Group 1 had the second greatest gain (4.11), while their counterparts in Group 2 had the smallest gain (1.64).



Note. G1 = Group 1; G2 = Group 2.

Figure 5. The relationship between treatment group and proficiency for textually explicit reading questions.

Proficiency Effects on Inferential Reading Questions

In order to examine the effect of proficiency (high, intermediate, and low) on the total scores of IF reading questions, the descriptive statistics for performance on IF reading question were produced (see Table 19).

The ANOVA summary displayed in Table 14 shows a statistically significant main effect for proficiency, $F(2, 63) = 14.756, p < .017$. The strength of association was estimated using eta squared. The main effect of proficiency accounted for 32% of the variance in performance on IF reading questions. Also, the interaction

between proficiency and treatment group was statistically significant, $F(2, 63) = 5.8989, p < .017$. The strength of association was also estimated using eta squared. The interactional effect accounted for 16% of the variance in performance on IF reading questions. However, there was a non-significant interaction between test administration time and proficiency. The statistically significant effects indicate

Table 19. *Descriptive Statistics for Inferential Reading Questions on the Pretest and Posttest for Three Proficiency Levels in Groups 1 and 2*

		Group 1			Group 2		
		Low <i>n</i> = 9	Inter. <i>n</i> = 12	High <i>n</i> = 13	Low <i>n</i> = 14	Inter. <i>n</i> = 11	High <i>n</i> = 10
Pretest	<i>M</i>	10.44	10.92	11.46	9.29	12.64	17.60
	<i>SE</i>	1.04	1.35	1.57	1.11	1.20	1.59
	95% CI						
	Lower Bound	8.04	7.94	8.04	6.90	9.96	14.01
	95% CI						
	Upper Bound	12.85	13.89	14.89	11.68	15.31	21.19
	<i>SD</i>	3.13	4.68	5.67	4.14	3.98	5.02
	Skewness	.26	-.60	.61	.69	-.58	.18
	<i>SES</i>	.72	.64	.62	.60	.66	.69
	Kurtosis	-.18	.33	-.42	-.19	-.35	-.94
	<i>SEK</i>	1.40	1.23	1.19	1.15	1.28	1.33
Posttest	<i>M</i>	12.33	15.25	15.54	10.21	14.18	19.50
	<i>SE</i>	.91	1.16	.97	.83	1.31	.78
	95% CI						
	Lower Bound	10.23	12.71	13.42	8.41	11.26	17.74
	95% CI						
	Upper Bound	14.44	17.79	17.66	12.01	17.11	21.26
	<i>SD</i>	2.74	4.00	3.50	3.12	4.36	2.46
	Skewness	1.34	-.29	-.67	.02	-.54	-.45
	<i>SES</i>	.72	.64	.62	.60	.66	.69
	Kurtosis	.96	-.51	-.55	.16	-1.18	-.42
	<i>SEK</i>	1.40	1.23	1.19	1.15	1.28	1.33

Note. Inter = Intermediate.

that the profile of performances across the participants with different proficiency levels was different for Groups 1 and 2.

In order to find the precise location of the proficiency differences, the estimated marginal means for the main effect of proficiency were calculated (see Table 20), and post hoc multiple comparisons of marginal means using the Bonferroni test were carried out (see Table 21). The result of the post hoc multiple comparisons showed statistically significant differences across all proficiency levels. The estimated marginal means were used to determine the nature of the interaction (Field, 2005) and are plotted in Figure 6. As Figure 6 shows, the score for Group 2 was higher for high proficiency participants than the score for Group 1; the scores for Groups 1 and 2 were very similar for intermediate proficiency participants; Group 1 scored slightly higher for low proficiency participants than Group 2. In general, this interaction seems to suggest that the performance of Group 2 on IF reading questions was more influenced by proficiency than the performance of Group 1. In other words, although the performance for Groups 1 and 2 fell as their proficiency level decreased, this fall was more pronounced for the participants in Group 2. A similar pattern of results was identified for IF reading questions just as for the TE and IF reading questions.

Table 20. *The Estimated Marginal Means for Proficiency Levels*

Group	Proficiency	Mean	SE	95% Confidence Interval	
				Lower Bound	Upper Bound
G1	High	13.50	.93	11.64	15.36
	Intermediate	13.08	.97	11.15	15.02
	Low	11.39	1.12	9.16	13.62
G2	High	18.55	1.06	16.43	20.67
	Intermediate	13.41	1.01	11.39	15.43
	Low	9.75	.90	7.96	11.54

Note. G1 = Group 1; G2 = Group 2.

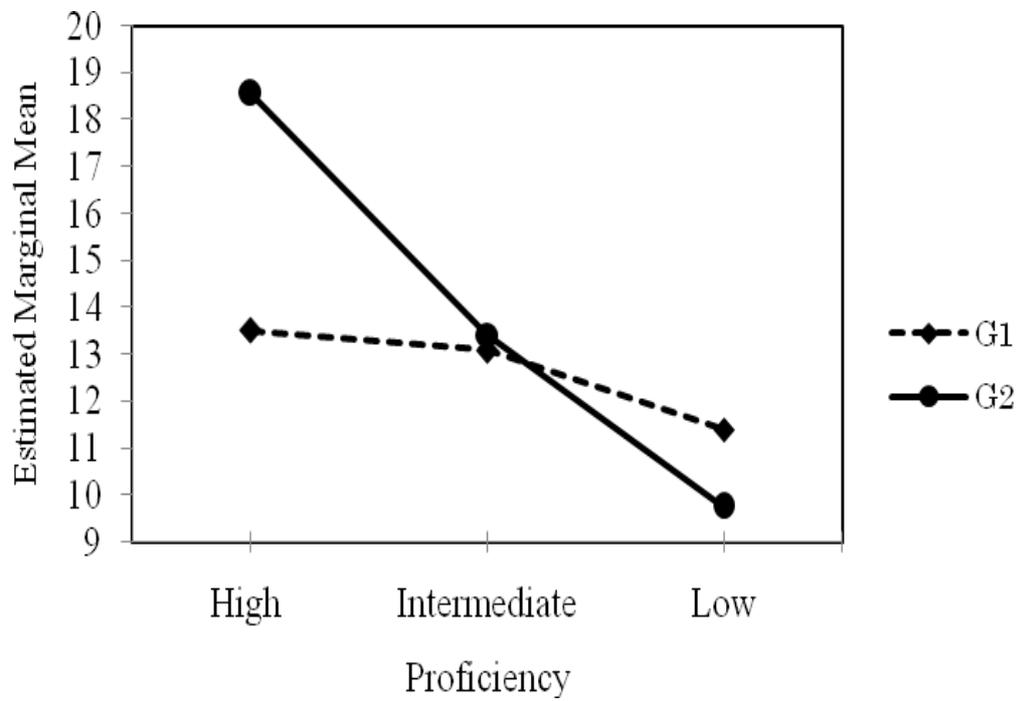
Table 21. *Bonferroni Test of Multiple Comparisons of Proficiency Levels*

(I) prof	J) prof	Mean Difference		p	95% Confidence Interval for Difference	
		(I - J)	SE		Lower-Bound	Upper-Bound
1	2	2.78*	.98	.020	.34	5.22
1	3	5.46*	1.00	.000	2.99	7.93
2	3	2.68*	1.00	.029	.22	5.14

Note. Based on estimated marginal means. * The mean difference is significant at

the .05 level. Adjustment for multiple comparisons: Bonferroni. 1 = high

proficiency level; 2 = intermediate proficiency level; 3 = low proficiency level.



Note. G1 = Group 1; G2 = Group 2.

Figure 6. The interaction between treatment group and proficiency for inferential reading questions.

CHAPTER 5

DISCUSSION

In this chapter, I will answer the research questions by interpreting the findings for each analysis and comparing them with the findings in previous studies.

Findings on Treatment Effects

The first research questions asked to what degree the test scores would differ when students received instruction emphasizing IF questions or instruction emphasizing TE questions. That is, did the teaching methods influence how the participants perform on the two types of reading questions? In order to determine whether any differences existed between Group 1, which received instruction emphasizing IF reading questions and Group 2, which received instruction emphasizing TE reading questions, I will examine the findings for analyses of performance in the following order on (a) both TE and IF reading questions, (b) only TE reading questions, and (c) only IF reading questions.

The findings for how the participants performed on TE and IF reading questions are as follows. First, a statistically significant main effect of test administration time (pretest and posttest) was found for the total scores of both TE and IF reading questions. This significant difference in mean scores for the pretest and posttest indicates that the participants in both groups benefited from their teaching methods with reading questions and therefore, performed significantly

better after the treatments than before. This result, in general, confirms a positive reading task consequence for how Japanese EFL learners read. It also demonstrates that because reading teachers in Japan are already familiar with the use of reading questions in class, they can continue using a task-based reading method with not only TE reading questions but also IF reading questions. TE reading questions are those that reading teachers have repeatedly exploited and can feel confident about due to their clearly presented answers. IF reading questions are those that some reading researchers have proposed introducing into reading classrooms, even when students are at the beginning level of proficiency (Day and Park, 2005; Grant, 1987; Nuttall, 2005; Tajika, 1999).

Second, there was neither a significant main effect for treatment group (different teaching methods) nor a significant interaction between test administration time and treatment group. These non-significant results show that although the teaching methods with reading questions significantly influenced how the participants performed, the way they performed did not differ greatly between Group 1 and Group 2. In other words, one teaching method did not have a statistically better effect on the way the participants read than the other. Nevertheless, statistical power to detect an effect was .20 for treatment group and .32 for the interaction. These observed values are far lower than the generally recommended value .80 for power (Cohen, 1988). Because of the low statistical power, these non-significant results must be cautiously interpreted. In other words, there might be a possibility of a Type II error (failing to observe a significant

difference when in truth there is one).

Added to the low statistical power, the descriptive statistics show that, in terms of the mean gain, Group 1 gained 7.38 points, while Group 2 gained 3.48 points (see Figure 1). The finding that the gains score of Group 1 was more than twice as much as that of Group 2 seems to be consistent with hypothesis 1: Gains will be greater for the group that receives instruction emphasizing IF questions than for the other group that receives TE questions because IF reading questions will encourage readers to fill in logical or pragmatic gaps during reading by making active use of background knowledge and inferences, and thus facilitate readers' cognitive involvement more with the text for meaning. This is because according to the cognitive view of reading ability that reflects the interactive nature of reading, text understanding results from the interactive integration of extracted text information and preexisting reader knowledge. The mean gain difference seems to imply that the activation of what is already known in the mind of a reader might work as a facilitative function to develop interactive integration and improve reading comprehension.

Next, I will present the findings for the analyses of participants' performance on TE reading questions. A similar pattern of results as identified in the previous analyses of the two types of questions was observed in TE reading questions. A statistically significant main effect for test administration time (pretest and posttest) was found, indicating that the participants in both Groups 1 and 2 benefitted from the task-based teaching methods and performed significantly better on TE reading

questions after the treatments than before as well. However, neither a significant main effect of treatment group nor a significant interaction effect between test administration time and treatment group was found. These non-significant findings indicate that one teaching method did not result in a significantly better effect on the participants' reading performance than the other. However, statistical power was .07 for the treatment group and .17 for the interaction. With the very low statistical power, these non-significant results must be interpreted very cautiously. There might be a possibility of a Type II error of failing to reject a null hypothesis when the alternative hypothesis is the correct one. In addition, the descriptive statistics show that Group 1 gained 3.79 points while Group 2 gained 2.09 points. As suggested in Figure 2, although Group 1 gained more and performed better than Group 2, the mean difference for the teaching methods did not reach a statistically significant degree.

The findings for performance on IF reading questions were similar to what was observed in the preceding two analyses. That is, a main significant factor effect for test administration time was found, but no treatment and interactional effects were indicated as significant. These results suggest that the participants in both groups benefited from their teaching methods and performed significantly better after the treatments than before, but one teaching method did not affect more beneficially how the participants read than the other to a statistically significant degree. Nonetheless, statistical power was .32 for treatment group and .42 for the interaction. These values are quite low for statistical power. Because of insufficient

statistical power, these non-significant results must be interpreted cautiously. As in the previous analyses, there is a possibility of a Type II error of accepting a hypothesis that should have been rejected. Furthermore, the descriptive statistics show that Group 1, which underperformed Group 2 on the pretest, outperformed Group 2 on the posttest by showing a more than twice higher gain (3.59) than Group 2 (1.40). This finding is similar to the one for the performance of both TE and IF reading questions.

In sum, the teaching methods with reading questions, whether primary emphasis was placed on TE or IF reading questions, made a statistically significant difference in how the participants read, but one teaching method did not have an influentially better effect on the participants' reading performance than the other.

However, the findings that Group 1 showed (a) almost the same gain scores on both types of reading questions (TE and IF reading questions), (b) higher gain scores on both types of reading questions than Group 2, and (c) more than twice higher overall gain scores than Group 2, seem to imply that the teaching method with primary emphasis on IF reading questions might have influenced the participants' reading more than the other method with main emphasis on TE reading questions. Also, as the gain difference for IF reading questions between Groups 1 and 2 was larger than for TE reading questions, instruction with primary emphasis on IF reading questions might have had greater influence on how the participants performed on IF reading questions. This finding would hardly be surprising in light of Day and Park's (2005) argument that students tended to

perform well on the types of comprehension questions that had been repeatedly employed by their teachers. A notable finding, however, is that Group 1 showed more or less the same gain scores for not only IF but also TE reading questions. Group 1's equally better performance on both types of reading questions seems to suggest that a teaching method emphasizing IF reading questions might facilitate lower-level processes (or literal text comprehension) as well as higher-level processes such as background knowledge use and inferencing. This finding seems to support the claims that Tajika (1999) put forth in her research about Japanese EFL readers' inferencing and that Day and Park (2005) made in their study on reading and their observations of reading classrooms. Tajika (1999) claimed that in order to develop better reading skills Japanese EFL learners should acquire more inference-oriented reading skills than decoding-oriented ones. Day and Park (2005) also suggested that more higher-order questions should be asked to encourage readers to go beyond a literal understanding of a text and to promote an interaction with the text and better understanding of what they read.

One reason for the statistically non-significant group differences observed in this study could be due to a lack of statistical power. The statistical power for group differences varied from .07 to .32 for the three ANOVAs. All the values were quite low. One factor associated with the insufficient statistical power might be a low precision of measurement achieved in this study. Since there were only thirty comprehension questions (15 TE and 15 IF reading questions with a full score of 60 points) on the short-answer test, the number of comprehension question items

might have been so small that Group 1's better performance could not be appropriately reflected in their scores, and as a result, the performance for Group 1 did not differ from that for Group 2 to a statistically significant degree. Another possible factor for the lack of statistical power might have been the sample size for this study: the total number of 69 participants (34 for one group and 35 for the other group) was very small. If few participants are used, a hypothesis test might result in such low power that there is little chance to detect a meaningful effect. Thus, the small sample size might be partly the cause of the non-significant statistical treatment effect. Increasing the validity of the measurement and the size of the sample might enable statistical power to detect a significant effect for group differences.

Findings on Proficiency Effects

The second research question asked to what degree the test scores would differ for high, middle, and low proficiency participants. That is, what effect would proficiency make on how the participants performed on the two types of reading questions in relation to two different treatments? Following the same procedures used in the preceding section for the treatment effects, I will report on the findings for the analyses of proficiency effects on how the participants performed on (a) both TE and IF reading questions, (b) only TE reading questions, and (c) only IF reading questions in this order.

First, I will report on the findings for the analyses of the participants' overall performance on TE and IF reading questions. The results show a significant proficiency effect but no significant interactional effects (see Table 8). The post hoc multiple comparisons of proficiency levels (high, intermediate and low) revealed two significant differences for proficiency: one between the high and low proficiency levels and the other between the intermediate and low proficiency levels (see Table 17). As suggested in Figure 4, there was a rather large and steep decline in estimated marginal mean scores between the three proficiency levels for Group 2. In contrast, there was a small and flat decline between the three proficiency levels for Group 1. Based on the interpretation of the nature of the major factor effect of proficiency in terms of the estimated marginal means (Field, 2005) for the two groups, it can be concluded that the statistically significant proficiency effects are due to the significant differences between the two higher proficiency levels and the low level for Group 2. This is because although the performances for both Groups 1 and 2 declined as the proficiency level decreased, this decrease was obviously more pronounced in Group 2. This finding for proficiency effects implies that Group 2 was more influenced by their proficiency levels than Group 1. In particular, proficiency made significant differences between the higher proficiency levels and the low proficiency level for Group 2.

On the other hand, for Group 1, proficiency made no significant difference. The findings on significant treatment effects and non-significant proficiency effects for Group 1 may lead to the conclusion that the participants in Group 1 benefited

from their teaching method (emphasizing IF reading questions) whichever proficiency level they belonged to. The descriptive statistics suggest evidence for this conclusion (see Table 15). According to Table 15, all the proficiency groups for Group 1 showed greater mean gains than their counterparts for Group 2 respectively. In Group 1, the high proficiency participants gained 8.46 points, the intermediate proficiency participants gained 7.25 points, and the low proficiency participants gained 6.00 points. The higher the proficiency levels they belong to, the more mean points they achieved. This finding concerning the mean gains for Group 1 is consistent with hypothesis 2, which stated that gain will be the greatest for participants at the high proficiency level within the group that receive instruction emphasizing IF reading questions because IF questions will match best the way that proficient readers read a text locally and globally for meaning, using bottom-up and top-down processing interactively.

Unlike Group 1, in Group 2 the high and intermediate proficiency participants had more or less the same mean gain (3.70 points for high and 4.46 for intermediate). The low proficiency participants gained 2.57 points. The finding for the smallest mean gain for the low participants in Group 2 is not consistent with hypothesis 3: Gains will be the greatest for participants at the lowest proficiency level within the group that receive instruction emphasizing TE questions because TE questions will match best the way unskilled readers read locally for locating and comprehending discrete pieces of text information, largely depending on bottom-up processing. The observed poor performance for the low proficiency participants in

comparison with the higher proficiency participants might have resulted from their limited language competence, namely, the linguistic threshold (Bernhardt & Kamil, 1995; Yoshida, 2007) or limited language sophistication (Koda, 2004). When poor readers face texts and assigned reading tasks that are too difficult for them, they are likely to resort to a slow and careful mechanical translation processing or to attempt to construct a situation model of the reader's interpretation from past experiences to force the text to match their preconceived notions (Ehara, 1996; Grabe & Stoller, 2002). In the former case, due to a lack of rapid and fluent bottom-up processing abilities such as lexical access, syntactic parsing and semantic propositional formation, working memory cannot function properly. In the absence of the proper operation of working memory, attention predominantly required for decoding detracts from what otherwise would be used for comprehension. In the latter case, a decontextualized situation model is imposed on reading comprehension by activating inappropriate background information and making wild guesses based on the inappropriately activated preexisting knowledge. For either case, their reading will result in poor comprehension.

A task-related strategy regarded as one of the comprehension variables (Enright, et al., 2000) might be another possible factor discriminating the higher proficiency participants from the lower proficiency participants (Cohen & Upton, 2006). The higher proficiency participants may be in a better position to make strategic decisions on where to go depending on the types of reading questions (Mark Sawyer, personal communication, February 19, 2008). That is, if they

understand that a posed question is text-bound, they search the text for relevant information in order to find answers to the textual question. If they decide that it is inferential, they search several parts of the text and integrate related pieces of information for making inferences. In this way the higher proficiency participants may be able to adapt their reading strategies flexibly depending on whether reading questions are literal or inferential because their cognitive resources are available for higher-order processing thanks to automaticity of bottom-up processing.

Another possible factor for why the low proficiency participants in Group 2 performed poorly might have something to do with an absence of explicit instruction about the use of cognitive reading resources to cope with text. Koda (2004) says that successful comprehension depends on both linguistic knowledge and the skills or strategies to utilize the knowledge for text-meaning construction. Therefore, if low proficiency students have enough language knowledge to pass a linguistic threshold but are not aware of what is cognitively available to them without explicit instruction accompanied by sufficient practice, they may not be able to make use of their cognitive reading resources adequately, particularly when they meet difficult text and reading tasks.

Next, I will report on the findings for the analyses of the participants' performance on TE reading questions. The results show neither a significant main effect for proficiency nor any interactional effects (see Table 11). However, the statistical power was .683 for proficiency, .051 for the interaction effect between test administration time and proficiency and .213 for the other interaction effect

between proficiency and treatment group. All these observed power values were not up to the benchmark value of .80 for power, so the non-significant effect results must be interpreted cautiously. In other words, there might be a possibility of a Type II error of accepting a hypothesis that should have been rejected.

One possible reason for the lack of a significant proficiency effect might be that the participants for this study were proficient at reading to locate and comprehend discrete pieces of information, which is a major basic skill required to answer TE reading questions. At the same time, directly relevant to this negative result is a lexical overlap with both question wording and correct answer (Nuttall, 2005; Raphael, 1982). Readers can answer text-bound questions mostly using the words in the text. The degree to which the correct answer matches the wording of the information in the text is regarded as one of the important task variables (Enright, et al., 2000) or test-taking strategies (Cohen & Upton, 2006). As the descriptive statistics (see Table 18) show, even the low proficiency participants in Group 2, who showed the lowest mean score on the posttest among all the proficiency groups, answered TE reading questions over 75% correctly (22.6 points out of 30). Moreover, the low proficiency participants in Group 1, who achieved the lowest mean score among the three proficiency levels in Group 1, answered more than 82% correctly (24.67 points out of 30). The developed basic reading skills of the participants and the task-based variable of the wording match might have resulted in the non-significant proficiency effect.

According to the descriptive statistics (see Table 18), the intermediate proficiency participants in Groups 1 and 2 performed almost in the same way on both tests and attained the exact same gains (2.91). In Group 1, the high and low proficiency participants showed almost the same mean gain points (4.38 for the high proficiency participants and 4.11 for the low proficiency participants), which was greater than the intermediate proficiency participants. Similarly in Group 2, the high and low proficiency participants showed almost the same mean gain points (1.80 for the high proficiency participants and 1.64 for the low proficiency participants), which contrastively was smaller than the intermediate proficiency participants. The contrast between Groups 1 and 2 in terms of the mean gains brings into prominence a wider gap of the mean difference between the two higher proficiency levels and the low proficiency level in Group 2 (see Figure 5). This finding partly seems to have something to do with the observed significant difference between the higher and low proficiency levels for Group 2 in the analyses of the overall performance on TE and IF reading questions.

Lastly, the participants' performance on IF reading questions will be reported. The results show a significant proficiency effect and a significant interaction effect between proficiency and treatment group (see Table 14). The post hoc multiple comparisons of proficiency levels revealed statistically significant differences across all proficiency levels (see Table 21). As suggested by Figure 6, Group 2 performed better than Group 1 for the high proficiency level, Groups 1 and 2 performed similarly for the intermediate proficiency level, and Group 1 performed

better than Group 2 for the low proficiency level. Based on the interpretation of the nature of the major factor effect of proficiency in terms of the estimated marginal means (Field, 2005) for the two groups, the estimated marginal means between the three proficiency levels for Group 1 was a rather small and flat decline. On the other hand, the estimated marginal means between the three levels for Group 2 was a rather large and steep decline. Although the performance of both Groups 1 and 2 declined as their proficiency level decreased, this decrease was more pronounced for Group 2. These findings suggest that the participants in Group 2 were more influenced by their proficiency levels than those in Group 1. The findings on proficiency effects for Group 2 are similar to what was identified for the participants' performance on the TE and IF reading questions. In regard to Group 1, the participants benefitted from their teaching method (emphasizing IF reading questions), but proficiency did not influence how they performed on IF reading questions. In other words, the participants benefited from their teaching method regardless of which proficiency level they belonged to. On the other hand, the participants in Group 2 benefited from their teaching method (emphasizing TE reading questions), but proficiency affected how they performed on the IF reading questions. This means that the participants in Group 2 benefited from their treatment, but the higher the proficiency the participants had, the more they benefited from the treatment.

The following reasons are postulated as possible causes for the proficiency effects on Group 2. The high proficiency participants who were already proficient

readers were able to interact with the text for meaning even though the small number of IF questions was given in order to stimulate their engagement in interaction with the text for meaning. The intermediate proficiency participants might have been proficient at searching the text in order to locate and comprehend discrete pieces of text information, largely depending on bottom-up processing. Also, they might have been somewhat on the way from a literal and local comprehension to interpretive comprehension where relevant intersentential information is integrated and, further, related to their background knowledge for making inferences. Even one IF question in the teaching method might have had a facilitative effect on their interpretive comprehension using bottom-up and top-down processing for constructing meaning. The low proficiency participants, due to limited coping strategies when they met difficult reading tasks like IF questions, due to a lack of explicit instruction (Koda, 2004) and repeated practice, might have had no choice but to resort to a slow mechanical translation processing (Grabe & Stoller, 2002). As a consequence, they could neither combine information from more than one source nor link combined information with background knowledge for inferencing. Or it might have been likely that although they might arrive at answers to IF reading questions, they might not have been proficient enough to express their answers (ideas or opinions) to IF reading questions. This is because IF reading questions mostly require them to use their own words rather than the wording of the text (Enright, et al., 2000).

The finding that the lower proficiency participants in Group 1 performed statistically better than their counterparts in Group 2 seems to support the claim that Nuttall (2005) and Grant (1987) made. That is, even beginning learners can answer reading questions that require them to go beyond a literal comprehension, and therefore the skills to read between the lines and beyond the lines should be introduced into a reading curriculum even at the beginner's level.

The descriptive statistics (see Table 19) show that all the proficiency participants in Group 1 showed a more than twice greater mean gain than their counterparts in Group 2 respectively. In each group, the higher proficiency participants gained more than the low proficiency participants. In Group 1, the high and intermediate proficiency participants showed a similar mean gain (4.08 points for the high proficiency participants and 4.33 points for the intermediate proficiency participants), and low proficiency participants gained a smaller mean gain of 1.89 points. Similarly in Group 2, the high and intermediate proficiency participants had more or less the same mean gain (1.90 points for the high proficiency participants and 1.54 points for the intermediate proficiency participants), and the low proficiency participants achieved a smaller gain of 0.92 points.

In sum, the findings on proficiency effects show that the participants in Group 1 benefited from their teaching method (emphasizing IF reading questions) in terms of performance on reading questions whichever proficiency level they belonged to. When it comes to the mean gain, the high proficiency participants

achieved the greatest gain in the total scores of the short test in Group 1 because IF questions may match best the way that proficient readers read a text locally and globally for meaning, using bottom-up and top-down processing interactively. Also, gain for all proficiency participants in Group 1 were higher than for their counterparts in Group 2 (a) on the total scores of TE and IF reading questions, (b) noticeably even on the scores of TE reading questions, and (c) on the scores of IF reading questions, except for the total scores of TE reading questions where the intermediate proficiency participants in the two groups showed the same mean gain. Therefore, the finding on the mean gain seems to imply that instruction emphasizing IF reading questions might have a more influential effect on how learners read. That is, this type of instruction might facilitate learners' processing of what they read.

Group 2 was more influenced by proficiency than Group 1 particularly in terms of performance on (a) the total scores of both TE and IF reading questions and (b) the total scores of IF reading questions. The higher proficiency participants benefitted more from their teaching method (emphasizing TE reading questions) than the low proficiency participants. Indeed, the low participants were proficient at reading to locate and comprehend discrete pieces of information, and/or they could correctly answer text-bound questions like TE reading questions correctly in the wording of the text. However, when it came to IF reading questions, the low proficiency participants performed poorly in comparison with the higher proficiency participants. Because of their limited language competence that does

not allow rapid and fluent bottom-up processing, they tended to resort to a slow and careful mechanical translation processing or to impose a decontextualized situation model on reading comprehension. Or they may not have been proficient enough to answer in their own words even though they were able to get the correct answers to IF reading questions. Due to the lack of explicit instruction about the use of cognitive reading resources to cope with text accompanied by sufficient practice, they might not have been able to utilize a task-related strategy to decide where to go to look for answers depending on the kinds of reading questions.

CHAPTER 6

CONCLUSION

In the last chapter, I will briefly summarize the main findings of the study, discuss its limitations, and make suggestions for future research.

Summary of the Findings

This study was an investigation of which type of reading questions, textually explicit or inferential reading questions, would best facilitate text comprehension, and how those reading tasks related to the language proficiency of Japanese EFL high school students.

The results of the current study indicated that teaching methods with reading questions, whether primary emphasis was placed on TE or IF reading questions, significantly influenced the reading comprehension of the participants. This finding shows the significant task effect of reading questions on how Japanese EFL learners read. It also demonstrates that textbooks and reading teachers can continue to use not only TE reading questions that have been predominantly designed and utilized (Applegate, et al., 2002; Ehara, 1996; Enright, et al., 2000; Nuttall, 2005; Thomas, 1992), but also should introduce a different kind of reading question, that is, IF reading questions, more often in the reading classroom (Day and Park, 2005; Grant, 1987; Nuttall, 2005; Tajika, 1999). However, one teaching method did not have a significantly better effect on the Japanese EFL learners' reading than the other.

Despite the lack of significant instructional effects in the quantitative analyses, which may be partly due to quite low statistical power varying from .07 to .32, Group 1 (that received instruction emphasizing IF reading questions) achieved (a) higher gain scores on both TE and IF reading questions than Group 2 (that received instruction emphasizing TE reading questions) and (b) more than twice higher overall gain scores than Group 2. These findings concerning the between-groups gain difference seem to imply that instruction emphasizing IF reading questions might have influenced the participants' reading more than instruction emphasizing TE reading questions. Furthermore, it is noteworthy that Group 1 achieved almost the same gain scores on not only IF reading questions, which the group had mainly been given, but also TE reading questions, which the group had not been focused on. This result may be because IF reading questions require readers first to scan the text locally in search of relevant surface information, and then, based on their integrated textual information, to fill in conceptual or pragmatic gaps between sentences or paragraphs. This process of searching for information on the way to working out answers to IF reading questions may have had a facilitative effect on their understanding of the textual information. This result might, therefore, suggest that instruction with primary emphasis on IF reading questions encourages Japanese EFL learners to understand what is explicitly stated as well as what is implied in the text during reading.

This study also showed that participants in Group 1 benefited from their teaching method regardless of proficiency level. In terms of gain, however, the

higher proficiency participants showed more mean gain scores, particularly in IF reading questions, than the lower proficiency participants. This may be because IF reading questions better match the way proficient readers read a text locally and globally for meaning, using bottom-up and top-down processing interactively.

In contrast, significant proficiency effects were found for Group 2. That is, participants in Group 2 were more affected by proficiency, particularly in terms of the total scores of (a) both TE and IF reading questions and (b) only IF reading questions, than those in Group 1. These results show that the higher proficiency participants benefited significantly more from their teaching method (emphasizing TE reading questions) than the lower proficiency participants. When it comes to TE reading questions, proficiency effects were not found significant because the low proficiency participants managed to locate and comprehend discrete pieces of information and/or they were able to answer text-bound questions mostly in the wording of the text. With regard to IF reading questions, however, the low proficiency participants performed poorly in comparison with the higher proficiency participants. This might be because when they met texts or assigned reading tasks that were too difficult for them, they seemed to have no choice but to resort to a slow and careful translation strategy or force the text to match their preconceived notions (Grabe & Stoller, 2002). Or due to their limited language proficiency, they might not have been able to express their answers to IF reading questions accurately using their own words even if they have had perceived the answers in their minds. Another possible reason is that due to a lack of explicit instruction about the availability of

cognitive reading resources to cope with text, essentially accompanied by sufficient repeated practice (Koda, 2004), they might not be able to use a task-related strategy that can guide them where to go to look for answers according to the kinds of reading questions.

Limitations

Although this study provides an improved understanding of reading questions and their effects on text comprehension of Japanese EFL learners, there are several limitations.

First, the statistical power of the data may not have been strong enough to observe quantitative effects of the group differences (the method differences). The statistical power for the group differences varied from .07 to .32 for the three ANOVAs. All the values were quite low against the standard value of .80. One possible factor for the insufficient statistical power was a low precision of measurement for this study. Because there were only thirty comprehension questions (15 TE and 15 IF reading questions with a full score of 60 points) on the short-answer test, the number of question items might have been so small that one group's better performance might not have been appropriately reflected in their scores. Also, the marking guidelines that the raters for the study followed might have had something to do with this internal validity problem. Raters took one point off from responses with the required information for not being written in acceptable language. This method of rating addresses language instead of understanding. In

particular, the responses to IF reading questions that low proficiency participants made might have been influenced by this rating system. Another possible factor for the lack of statistical power was the sample size for this study. This study had a total number of 69 participants (34 for one group and 35 for the other group). If few participants are used, a hypothesis test could result in such a low power that there is little chance to detect a meaningful effect. Thus, the small sample size might be partly the cause of the non-significant statistical treatment effect. Increasing the validity of the measurement and the size of the sample might make the statistical power strong enough to detect a significant effect for group differences.

Second, the treatment period for this study was five months. This time length roughly amounts to one semester in Japanese high schools, which have adopted a two-semester system. If the treatment had been extended to one academic year, Japanese EFL learners might have been more noticeably influenced by their instruction with reading questions and the effect of instruction might have been more clearly indicated.

Third, although two types of reading questions, TE and IF reading questions, were investigated in the current study, IF reading questions can be subdivided into two types according to the information source that is accessed for inferencing: text-based inferences or knowledge-based (or schema-based) inferences. Inferencing in text-based inferential questions depends heavily on textual information. Readers are required to draw text-based inferences logically to fill in conceptual or pragmatic gaps, based on the semantic content of explicitly stated propositions in

the text. Text-based questions whose answers are hidden in relationships between explicitly stated information in single sentences or small portions of the text are labeled Textually Implicit Questions (Pearson & Johnson, 1978) or Search-and-Think (Raphael, 1982, 1986). In contrast, inferencing in knowledge-based inferential questions depends primarily on the readers' background knowledge. Readers are required to draw schema-based inferences logically to fill in conceptual and pragmatic gaps, based on their beliefs. These knowledge-based questions whose answers are based on overall understanding of the ideas and organization of the complete text passage but largely reside in the readers' minds are labeled Scriptally Implicit Questions (Pearson & Johnson, 1978) or On My Own (Raphael, 1982, 1986). If three groups of participants had been included in the study and assigned to instruction emphasizing TE reading questions, text-based IF reading questions or schema-based IF reading questions, more detailed findings concerning the effect of reading questions on the comprehension of the Japanese EFL learners might have been obtained.

Finally, a more reasonable research design would have used three classes instead of two: one class that received instruction emphasizing TE reading questions, another class that received instruction emphasizing IF reading questions, another class that received instruction using the Grammar-Translation method. The three-group design would make it possible to compare a task-based teaching method with reading questions and the traditional grammar translation method that

is still used by many Japanese reading teachers in addition to teaching methods with different types of reading questions.

Suggestions for Future Research

Participants in the study were Japanese public senior high school juniors, aged 16-17, studying in the same school. This type of study should also be conducted with public junior high school students, who are just starting to take English lessons and learning how to read using instructional material. Those students are, in the true sense of the meaning, beginning EFL learners. An investigation of the effect of the two types of reading questions on the text comprehension of Japanese beginning EFL learners could be compared with what was found in both this study and the study that was carried out by Day and Park (2005).

The reading task employed for reading instruction in this study included two types of reading questions, textually explicit and inferential reading questions. The study showed that a teaching method with primary emphasis on IF reading questions might influence text comprehension of the Japanese EFL learners more than one with primary emphasis on TE reading questions. That is, teacher-devised inferential reading questions might facilitate the learners' interaction with text for meaning. In order to further activate this interactive processing and promote their active contribution to creating text meaning, it is possible to swap one of the teacher-devised inferential reading questions with a learner-devised one. That is,

the learners are encouraged not only to fill in conceptual or pragmatic gaps in the text based on what they have understood but also to use their inferencing skills to find other logical gaps in the text. The strategy to ask learners to make questions has been used in inference studies with think-aloud protocols (Kadota & Noro, 2001). The same strategy employed for the think-aloud protocols could be applied to reading instruction and its effects compared with the findings of this study.

Final Conclusions

An appropriate understanding of reading question types and their effects on text comprehension of Japanese EFL learners will make it possible for reading teachers to make important changes in their teaching practices and encourage Japanese EFL learners to move from being passive text decoders to active meaning makers. Reading teachers should use a task-based teaching method with reading questions. If the use of reading questions is already a part of reading teachers' methodology, they should include not only commonly used textually explicit reading questions but also inferential ones. Implementing these changes will help break the cycle of translation-bound reading instruction with its overemphasis on lower-level processing, and will lead students to read text in a more meaningful, interactive way.

REFERENCES CITED

- Aebersold, J. O., & Field, M. L. (1997). *From reader to reading teacher*. New York: Cambridge University Press.
- Applegate, M. D., Quinn, K. B., & Applegate, A. J. (2002). Levels of thinking required by comprehension questions in informal reading inventories. *The Reading Teacher*, 56(2), 174-180.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bernhardt, E.B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, 16, 15-34.
- Birch, M. B. (2007). *English L2 reading: Getting to the bottom* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, New Jersey: Prentice Hall Regents.
- Brown, J. D. (1997). Questions and answers about language testing statistics: skewness and kurtosis. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 1(1), 18-20.
- Carrell, P. L. (1988) SLA and classroom instruction: Reading. *Annual Review of Applied Linguistics*, 9, 223-242.
- Champeau De Lopez, C. L., Marchi, B. G., & Arreaza-Coyle, M. E. (1997). A taxonomy: Evaluating reading comprehension in EFL, *FORUM*, 35(2), 30-37.
- Chikalanga, I. (1991). *Inferencing in reading: A cross-cultural study*. Unpublished doctoral dissertation, University of Reading.
- Chikalanga, I. (1992). A suggested taxonomy of inferences for the reading teacher. *Reading in a Foreign Language*, 2, 697-709.
- Chikalanga, I. (1993). Exploring inferencing ability of ESL readers. *Reading in a Foreign Language* 10(1), 931-952.

- Clark, M., & Zhang, Y. (2005). *Broadening international student access: Looking at alternative English proficiency tests*. Tokyo: The Society for Testing English Proficiency.
- Cohen, A. D., & Upton, T. A. (2006). *Strategies in responding to the new TOEFL reading tasks*. (TOEFL Monograph Series No. MS-33). Princeton, NJ: ETS
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum.
- Cunningham, R. T. (1971). Developing question-asking skills. In J. Weigand (Ed.), *Developing teacher competencies*. Englewood Cliffs, NJ: Prentice Hall.
- Day, R. R., & Park, J. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, 17(1), 60-73.
- Developing oral communication abilities: Promoting the unification of English education in high schools. (2007, October 7). *The Asahi Newspaper*, p. 31.
- Ehara, K. (1996). Reading in a foreign language: The effects of cognition-activating questions compared to fact-seeking questions for reading activities. *Studies on Language*, 5. Tokyo: Tokyo Applied Linguistic Institution.
- Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M., (2000). *TOEFL 2000 reading framework: A working paper* (TOEFL Monograph No. MS-17). Princeton, NJ: ETS
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.
- Grabe, W., & Stoller, F.L. (2002). *Teaching and researching reading*. Essex, UK: Longman.
- Graesser, A.C., & Kreuz, R.J. (1993). A theory of inference generation during text comprehension. *Discourse Processes*, 16, 145-160.
- Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- Grant, N. (1987). *Making the most of your textbook*. London: Longman.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House Publishers.

- Heaton, J. B. (1988). *Writing English language tests*. (New ed.) London: Longman.
- Heilman, A., Blair, T., & Rupley, W. (1986). *Principles and practices of teaching reading* (6th ed.). Columbus: Merrill.
- JACET SLA research group. (2005). *Bunken kara miru dainigengo shutoku kenkyu (Literature search in second language learning research)*. Tokyo: Kenkyusha.
- Kadota, S., & Noro, T. (2001). *How the mind works in EFL reading*. Tokyo: Kuroshio Publishing Company.
- Kanatani, K. (1995). *Eigo readingron (English reading theories)*. Tokyo: Kagensha.
- Kawasaki, A. (1998). *Survey of high school English teachers in Fukui Prefecture. In Japanese and English rhetorical strategies: A contrastive analysis with pedagogical implication*. Unpublished graduation thesis: Fukui University.
- Kembo, J. A. (2001). Testing of inferencing behavior in a second language. *International Journal of Bilingual Education and Bilingualism*, 4(2), 77-96.
- Koda, K. (2004). *Insights into second language reading*. New York: Cambridge University Press.
- Kosuge, K. (1991). *Yakudokushiki no jyugyo wo doukoeru?* (how should grammar translation method be conquered?). *Modern English Teaching*, 4, 16-17.
- Monbusho (1999). *Kotogakkou gakushuushidoyoryo kaisetsu (guidelines for the course of study for upper-secondary school: Foreign languages)*. Tokyo: Kairyudo.
- Mulvey, B. (1998). Entrance exams: The reading example. *ONCUE*, 6(3), 5-12.
- Mulvey, B. (1999). A myth of Influence: Japanese university entrance exams and their effect on junior and senior high school reading pedagogy. *JALT Journal*, 21(1). 125-142.
- Muranoi, H. (2006). *SLA research and second language learning and teaching*. Tokyo: Taishukan.
- Nicholas, D. W., & Trabasso, T. (1980). Toward a taxonomy of inferences for story comprehension. In Wilkening, J. B., & Trabasso, T. (Eds.), *Information integration by children* (pp. 243-265).

- Nuttall, C. (2005). *Teaching reading skills in a foreign language*. Oxford: Macmillan.
- The STEP Test in Practical English Proficiency. (2006). *Eiken nikyu zenmondaishu (the step test in practical English proficiency for 2nd grade)*. Tokyo: Obunsha.
- Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.
- Raphael, T. E. (1982). Question-answering strategies for children. *The Reading Teacher*, 36, 186-191.
- Raphael, T. E. (1986). Teaching question-answer relationships, revised. *The Reading Teacher*, 39, 516-522.
- Saito, S. (1992). *Schema riron to dokkainoryoku (schema theory and reading comprehension ability)*. *The Language Teacher*, 41(10), 14-16.
- Sano, M., Matsumura, K., Umeda, O., Kawasaki, A., Nishizawa, M., & Snelling, J. (2004). *Genius English writing*. Tokyo: Taishukan.
- Shimizu, M. (2005). Reading test niokeru shitsumon type (types of questions used on reading comprehension tests). *Step Bulletin*, 17, 48-62.
- Suzuki, H., Snowden, P., & Eto, S., Kaji, Y., Yamamoto, T., Ohashi, K., Suzuki, S., & Hakomori, T. (2003). *New legend English II*. Tokyo: Kaitakusha.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins College Publishers.
- Tajika, H. (1999). Elaborative inferences in reading comprehension. *The Annual Report of the Institute for Research in Language and Culture at Tsuda College*, 14, 128-142.
- Takanashi, M. & Takahashi, M. (1991). *Eigo reading shido no kiso* (4th ed.) (*the basics for instruction of English reading*). Tokyo: Kenkyusha Publishing Company.
- Tenma, M. (1992). Dokkaishido no houhouon: dokkai no strategy (theories for reading instruction: reading comprehension strategies). *The English Teachers' Magazine*, 41(10), 11-13.

- Tenma, M., Araki, K., Fukaya, K., Horiba, Y., Isahara, H., Konishi, K., Miyaura, K., Murasugi, K., Narita, M., Ozaki, K., & Tajika, H. (2002). *Eibundokkai no process to shido (processes and instruction of reading English as a foreign language)*. Tokyo: Taishukan.
- Terauchi, M. (1991). Koukou no dokkaishidou ni kansuru jittaicyosa (investigation into how reading is taught at senior high schools). *Perspective*, 87, 3-6.
- Thomas, B. B. (1992). How can we use what we know about questioning skills to develop literate thinkers? *Reading Horizon*, 33(1), 19-30.
- Tsuchiya, S. (1992). Koukou reading no mokuhyou to shidou (objectives and instruction for reading in senior high school). *The English Teachers Magazine*, 41(10), 8-10.
- Vacca, J. L., Vacca, R. T., & Gove, M. K. (1991). *Reading and learning to read* (2nd ed.). New York: Harper Collins Publishers.
- Yamashita, J. (2003). Process of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20 (3), 267-293.
- Yoshida, M. (2007). *The effects of task, text, and proficiency on second language reading*. Unpublished doctoral dissertation, Temple University.
- Warren, W. H., Nicholas, D. W., & Trabasso, T. (1979). Event chains and inferences in understanding narratives. In Freedle, R. O. (Ed.), *New directions in discourse processing* (pp. 23-52). Norwood, NJ: Ablex.

APPENDIX A

TAILORED CLOZE TESTS

Passage 1

Directions: Read the following story and fill in the blanks with proper words.

The world's largest population of wild cheetahs lives in the African country of Namibia. Over 90 percent of these cheetahs, however, live on land owned by farmers. Normally, the hunting of these endangered animals (1) not allowed by the Namibian government, but (2) can shoot them when they attack their (3) and goats. In order to (4) the cheetahs, while also helping farmers keep their animals (5), a conservationist organization has started the Livestock (6) Dog Program.

Namibian farmers already (7) a tradition of using dogs to keep their (8) together in one place. When a (9) approaches the farm animals, these dogs try to move them (10) a safe place. Unfortunately, this can actually encourage (11) to attack because cheetahs naturally (12) after running animals. In order to solve this (13), the Livestock Guarding Dog Program uses (14) called Anatolian shepherds instead of the local (15). These Anatolian shepherds are not trained to chase or attack---they just (16) loudly. Because cheetahs do not fight (17) other large animals, the barking is enough to scare (18) away.

Anatolian shepherds are well suited to the job. These (19) originally come from a dry, flat region in Turkey with a climate very similar (20) Namibian's. Their excellent eyesight and hearing, together with their strong instinct to protect, make them perfect guard dogs. Under the program, an Anatolian shepherd puppy is raised with a group of farm animals so that it gets to know them and learns to act as their protector.

The farmers involved with the program are pleased with their new dogs and find that their animals are being well protected. Conservationists are also happy that the number of cheetahs killed by farmers each year has greatly decreased. Now, a number of other countries have shown interest in using these guard dogs to protect farm animals as well as local endangered species.

Model Answers

1	is	2	farmers	3	sheep	4	protect
5	safe	6	Guarding	7	have	8	animals
9	cheetah	10	to	11	cheetahs	12	chase
13	problem	14	dogs	15	dogs	16	bark
17	with	18	them	19	dogs	20	to

Passage 2

Directions: Read the following story and fill in the blanks with proper words.

Many people think that sleeping for a short time during the day is only for small children or old people. However, some of the world's most famous people, such as Napoleon Bonaparte, Albert Einstein, and Winston Churchill, took regular naps when they were adults. They knew something that many people (1) discovering again today---the fact that (2) a nap in the afternoon greatly increases one's ability (3) work.

Researchers say that the (4) time for napping is about eight hours after waking (5) in the morning. In fact, the human body (6) naturally set to rest in the middle of the day. (7) is why people in a number of countries take (8) for a short sleep sometime after (9). Shorter naps---those under 30 minutes---(10) often better than longer naps. This (11) because when people sleep longer, they also sleep more deeply (12) then have difficulty waking up.

Companies that encourage the practice of (13) an afternoon nap are finding (14) it has a positive effect. Employees who nap work faster, make fewer (15), and have fewer accidents than (16) who do not. One study, for example, showed that airline (17) who took naps for about 25 minutes greatly improved (18) ability to concentrate.

Most companies still do (19) allow their workers to nap during (20) day. This is probably because few bosses want to pay people to sleep. But even though workers who take naps during business hours may look lazy, they are likely to end up achieving more for their companies.

Model Answers

1	are	2	taking	3	to	4	best
5	up	6	is	7	That	8	time
9	lunch	10	are	11	is	12	and
13	taking	14	that	15	mistakes	16	those
17	pilots	18	their	19	not	20	the

APPENDIX B

SHORT-ANSWER TESTS

Passage 1

Directions: Read the following story and answer the questions.

James Hogan, 38, became a father 22 months ago. When his daughter was born, he quit his full-time job. He wanted to have more time to spend with his child.

“I wanted to be a real father to my daughter,” says Hogan of Washington, D.C.

At first, Hogan had a lot of problems. “I had a lot to learn about babies,” he says. “For example, I didn’t know how to change a diaper and I wasn’t sure how to hold a baby.”

There are thousands of fathers like Hogan in the United States today. These “liberated dads” are spending more time with their children---holding them, feeding them, playing with them, and changing their diapers.

There are many reasons for this change. One reason is that modern men have more free time away from work. Another reason is that more married women with children work outside the home. These mothers are too busy to do all the housework and take care of the children alone. Their husbands have to help them.

Modern fathers are enjoying the change. James Hogan spends 25 to 30 hours per week with his daughter. He says he isn’t sorry he left his job. “I wanted more time to be with my daughter, and now I have it,” he says.

Another father who is enjoying the change is Harlan Swift, a 34-year-old engineer and father of two. “Spending time with the kids is the most important thing in my life. I enjoy every minute I’m with them,” Swift says.

Not many men can afford to quit work when they become fathers. They don’t have enough money. Combining a full-time job and home life is difficult. You need a lot of energy. As Harlan Swift says, “I never have enough time to do everything.”

Modern fathers will need to learn to balance work at home with their jobs outside the home. They will also need to learn new skills. Like James Hogan, many men are uncomfortable holding babies or changing diapers. That is not going to change overnight.

Questions

Answer all the questions in English. (すべて英語で答えなさい)

1. How long has James Hogan been a father?
2. How old was James Hogan when he became a father?
3. What did James Hogan do when his daughter was born?
4. Where does James Hogan come from?

5. What new skills did James Hogan need to learn after becoming a father?
6. What are thousands of fathers like James Hogan called?
7. Why do more married women with children work outside the home?
8. What does 'this change' on line 13 mean?
9. What does 'their' on line 17 refer to?
10. How long does James Hogan spend with his child every week?
11. What does 'it' on line 20 refer to?
12. What is Harlan Swift's job?
13. How many children does Harlan Swift have?
14. Why is combining a full-time job and home life difficult?
15. What will modern fathers need to learn?

Passage 2

Directions: Read the following story and answer the questions.

We usually think of bears as living in the wild, but these days more and more of them are spending time in towns. A study by the Wildlife Conservation Society of the United States has shown that the main cause of this is the garbage that is found near fast-food restaurants and people's homes. Garbage is a great source of food for bears because they can find it in the same place all year around.

The researchers followed 59 bears for 24 hours in the Lake Tahoe area between Nevada and California. They found that there were two types of bears: "country bears," who spent nearly all their time in the wild, and "city bears," who stayed mainly in the towns. Bears are usually active during the day, but the city bears rested during the day and were active at night. While country bears spent a lot of time searching for food, city bears simply ate the never-ending supply of food in people's garbage. The result is that some city bears have grown up to 270 kilograms ---twice the weight of normal bears.

Eating human food does not itself seem to harm the bears. Rather, the problem is that towns are very dangerous places for them. When bears come near humans, they often get killed. Most die in traffic accidents, but an increasing number of bears are killed in other ways. For example, when the bears break into people's homes, wildlife officials often have to shoot them for safety reasons.

In order to protect both the local community and the bears, a campaign called "A Fed Bear Is a Dead Bear" has been started. Campaigners say that the best way to protect bears is to stop them from eating garbage by using garbage containers that are too difficult for them to open. When people do this, the bears leave, and this makes life safer for both animals and human beings.

Questions

Answer all the questions in English. (すべて英語で答えなさい)

1. What places does 'the wild' on line 1 mean for bears?
2. In what country was the Wildlife Conservation Society's study carried out?

3. What does 'this' on line 4 refer to?
4. What kind of human food will bears find in the garbage?
5. Why is garbage a great source of food for bears?
6. Where are the researchers from?
7. How many types of bears did the researchers find in the study?
8. Which bears rested during the day?
9. What effect does a city life have on city bears?
10. About how much do normal bears weigh?
11. Why is it dangerous for bears to be in towns?
12. Write all the missing words after 'Most' on line 18.
13. What do wildlife officials use when they shoot the bears for safety reasons?
14. How can people stop bears from eating garbage?
15. What may be another possible reason for more bears to come down to towns?

Model answers for Passage 1

Note: The expressions in parentheses are optional information.

1. 22 months.
2. 36 or 37 (years old).
3. He quit his full-time job.
4. (He comes from) Washington, D. C.
5. (He needs to learn) how to change a diaper and how to hold a baby.
6. (They are called) "Liberated dads."
7. "Liberated dads"/Fathers are spending more time with their children.
8. Because of her life style/Because they need make money do support their family, etc.
9. Married women with children working outside.
10. (He spends) 25 to 30 hours (every week).
11. More time to be with his daughter.
12. An engineer.
13. Two (children).
14. Because you need a lot of energy and don't have enough time to do everything.
15. (They will need) to balance work at home with their jobs outside the home and to learn new skills.

Classification of reading questions and rationale for their selection

1. Textually explicit

The answer to this question is explicitly stated on the page. Readers are required to locate a discrete piece of information.

2. Textually explicit

The question requires readers to reinterpret Hogan's age, 38, and the paternal period, 22 months, involving a numerical calculation to arrive at the answer.

3. Textually explicit

The answer to this question is explicitly stated in the text. The question can be answered in the words of the text due to lexical overlap between the question and the expected information in the text. Readers are required to comprehend a sentence-level proposition.

4. Textually explicit

The question involves the recognition of information described in words that differ from those used in the text. That is, 'from' is used in the question while 'of' is found in the text. However, the answer to this question is explicitly stated in the text. Readers are required to search the text for a discrete piece of information, based on the lexical and grammatical clue 'of' introducing the place people belong to.

5. Inferential

The question requires readers first to infer from their knowledge of the context what new skills are intended: skills for caring for babies. Then, it requires readers to search for information related to baby-caring skills stated in the text by using their domain-specific background knowledge about baby-caring. Finally, it requires readers to integrate discrete pieces of relevant information from more than one source of the text to answer the question.

6. Textually explicit

The answer to the question can be taken verbatim from the text. Readers can arrive at the answer by recognizing the pro-form 'these' which indicates the preceding part of the text and a lexical cohesion synonymy between 'fathers' and 'dads.'

7. Inferential

The answer to this question is not stated in the text, so readers need to answer by making use of their relevant background knowledge about working mothers with children and providing a plausible reason for the given situation.

8. Inferential

The question requires readers first to understand that the pro-form 'this change' implies an occurrence of something new or different stated in the preceding part of the text. Then, readers need to retrieve its referent from the text by using their knowledge of the context and their schemata about fathers' roles at home.

9. Inferential

The question requires making two pronominal inferences in the process of identifying what the pronominal 'their' refers to. That is, readers are required to take two steps to arrive at the answer. The first step is to identify what 'their' refers to in the closest context (i.e. 'these mothers' in the preceding sentence). However, the referent remains ambiguous in its meaning because the pro-form 'these' in it does not clarify what these mothers are like. The second step required of readers is, therefore, to go further back to the preceding part of the text and locate what 'these mothers' refer to: 'married women with children who work outside the home.'

10. Textually explicit

The answer to the question is directly presented in the text. Readers are required to search for a discrete piece of information in the text.

11. Inferential

The question requires making a pronominal inference in order to identify what 'it' refers to. Readers are required to make coherent sense of the two sentences by inferring an implied contextual relationship between the pronominal 'it' and its referent 'more time to be with my daughter.' That is, it requires recognizing the functional values of the two sentences. The preceding sentence describes Harlan Swift's past wish, and the following sentence with 'it' describes the realization of the desire.

12. Textually explicit

The answer can be directly and precisely taken from the text. Readers are required to locate a discrete piece of information.

13. Textually explicit

The question has its obvious answer right there on the text. Readers are required to locate a discrete piece of information.

14. Inferential

The question requires readers to infer an implied contextual relationship between the target sentence and the following two sentences. In other words, readers need to recognize the functional values of the three sentences. The following two sentences after the target sentence can be interpreted as instantiating what is claimed in the target sentence. Readers need to search for bits of relevant information from two sources in the text and put them together to answer the question.

15. Textually explicit

The answer is clearly stated in the text. Readers can answer the question in the words of the text after they connect pieces of relevant information from the two sentences following the target sentence.

Model answers for Passage 2

Note: The expressions in parentheses are optional information.

1. In nature/woods/forests/mountains, etc.
2. (In the) United States.
3. More and more bears are spending time in town.
4. Bread/Hamburgers/Fish/Rice, etc.
5. Because they can find it (in the same place) all year around.
6. (They are) from the Wildlife Conservation Society of the U.S.
7. Two (types).
8. The city bears (did).
9. They rest during the day, become active at night, and get fatter.
10. (About) 135 kilograms.
11. Because they come near humans, they get killed/Because they can be killed in a variety of ways.
12. of the bears that often get killed
13. (With) guns/rifles.

14. (They can stop them from eating garbage) by using garbage containers that are too difficult for them to open.
15. (Because of) lack of food in the wild/disappearing woods, etc.

Classification of reading questions and rationale for their selection

1. Inferential

The answer is not found on the page but in the minds of the readers. The question requires readers to use their knowledge of the context concerning bears living in the wild and also, their relevant domain-specific knowledge about places inhabited by bears in order to arrive at the answer. In other words, the question requires inferring a subcategory or a particular exemplar that instantiates an explicit noun phrase 'the wild.'

2. Textually Explicit

The answer is explicitly stated in the text. Readers are required to find a discrete piece of information that represents a specific country. Or, if they recognize a grammatical relationship between the phrase 'the Wildlife Conservation Society' and the following prepositional phrase 'of the United States,' they can arrive at the answer.

3. Inferential

The question requires readers first to infer from their knowledge of the context concerning what the pro-form 'this' implies and then to identify its referent. That is, readers need to understand that 'this' will indicate an occurrence or event, based on its modified preceding phrase 'the main cause.' Next, readers have to go back to the preceding part of the text and search for what happened.

4. Inferential

The answer is not stated in the text, so readers are required to infer an example of human food included in the garbage using their knowledge of the context (in particular, 'near fast-food restaurants and people's homes') and relevant background knowledge. The question requires readers to infer a subcategory or a particular exemplar that instantiates the explicit noun phrase 'human food.'

5. Textually explicit

The answer is explicitly stated in the text. The recognition of a subordinate conjunction 'because' after the target sentence enables readers to arrive at the answer. In other words, the cause of garbage being a great source of food for bears is linguistically marked.

6. Inferential

The question requires readers to search the text for the organization or institution 'the researchers' belong to, making use of their relevant background knowledge about a research body and its members. Also, because there is no explicitly stated grammatical or lexical clue in the sentence with the target noun phrase in it, readers need to infer from the definite article 'the' preceding 'researcher' that their organization or institution is already mentioned in the preceding part of the text.

Based on their domain-specific background knowledge and recognition of the discourse marker, readers need to arrive at the answer.

7. Textually explicit

The answer is explicitly stated in the text. Readers are required to locate a discrete piece of information in the text.

8. Textually explicit

The answer is explicitly stated in the text. The question requires searching for a discrete piece of information in the text.

9. Inferential

The question requires readers to infer implied causal consequences of “city bears” staying in towns based on the appropriate understanding of the text. Readers need to integrate pieces of relevant information from various parts of the second paragraph and to organize them in the process of arriving at the answer.

10. Textually explicit

Readers are required to reinterpret the information ‘270 kilograms---twice the weight of normal bears’ by carrying out a numeral calculation.

11. Inferential

Readers need to infer the functional values of the target sentence with the key word ‘dangerous’ and the following sentence in terms of the contextual relationships between them. That is, in the process of arriving at the answer, readers have to understand that the target sentence functions as an assertion that towns are dangerous places for the bears and that the following sentence functions as a reason for the assertion.

12. Inferential

The question requires readers to fill in the grammatical form of the omission with information from the previous part of the text considering the structure of the target sentence. Readers need to grammatically recognize ‘most’ as a subject of the target sentence and also semantically as living things due to the following verb ‘die.’ So, readers will connect ‘most’ and ‘bears.’ At the same time, readers have to interpret the functional value of the target sentence in terms of its relationship with the preceding sentence. In this case, the previous sentence functions as a reason for the assertion that bears that come near humans often get killed. The functional value of the target sentence is to instantiate how they lose their lives. Based on the appropriate understanding of the implied contextual structure where a reason is followed by an example, readers are required to give the answer ‘most of the bears that often get killed.’

13. Inferential

The question requires readers to infer an instrument used when the agent ‘wildlife officials’ executes an intentional action. Because the answer is not presented in the text, readers need to fill in this pragmatic gap with information using their world knowledge and relevant background knowledge about weapons used to shoot wild animals.

14. Textually explicit

The answer is explicitly stated in the text. That is, a way of stopping bears from eating garbage is linguistically marked. Readers need to recognize the key phrase 'stop bears from eating garbage' appears in both the question and the text. Then, readers can answer using the following phrase verbatim taken from the text.

15. Inferential

Readers are required to infer what will cause more bears to come down to human living areas using their world knowledge and relevant background knowledge based on the appropriate understanding of the whole text.

APPENDIX C

THE COURSE SYLLABUS OF ENGLISH II

Unit	Title and its brief content	Topic	Text type	Grammar points	Key phrases
1	What Happened to Sparky? A story about a boy named Sparky, who created “Peanuts”	Autobiography	Expository	Past subjunctive, Past perfect passive, Relative pronoun	get along, be afraid of, turn ~ down, make up one’s mind, be proud of, give up
2	Hold That Call! Cellular phones and their bad effects on the human body	Science and health	Expository	Past perfect continuous, Negation of infinitive, Pronouns	be popular with, have an effect on, blame ~ on, give ~ out, keep ~ away from, carry ~ out
3	Laughter— Everyone’s Language Laughter and its good effects on the human body	Laughter and its effects	Expository	Nonrestrictive relative pronouns, Sentence subjects, Expressions showing concession	all around, join in, to oneself, put ~ up, run into
4	“Waste Not, Want Not” Classroom discussion about garbage problems	Environmental problem	Conversation	Relative adverb <i>why</i> , Subjunctive without using <i>if</i> , Auxiliary verbs in past tense	put ~ in danger, throw ~ away, these days, a ~ point of view, so far, another day

5	A Trip on the Trans-Siberian A slow trip between Tokyo and Moscow by sea and Trans-Siberian Express	Travel and culture	Travel sketch	Nonrestrictive relative adverbs <i>where</i> and <i>when</i> , <i>Will/shall +have+</i> past participle, To-infinitive	as far as, share ~ with, become friends with, as long as, take ~ along, this time
6	The Negative Sides of Fast Food A report about the consumption of fast food in the U.S. and bad effects of fast food on the human body	Society and health	Expository	Future progressive, Passive voice of infinitive, Subjects of infinitive clauses	the same with, according to, short of, be used to, tend to <i>do</i> , combine ~ with
7	Advertising-Selling a Product A report about advertisers' strategies to make consumers buy one brand instead of another	Psychology	Expository	S+V+O+O[= <i>if, whether, and wh-word</i>], Partial negation, Usage of a causative verb <i>get</i>	instead of, problem with, lead ~ to <i>do</i> , see ~ as, reach out
8	Long Walk to Forever A story about a soldier who returns to his home to confess his love to a girl who has been his friend since childhood	Love story	Narrative (love story)	As if/though + the past subjunctive, Collective nouns, Ellipsis	belong to, have no idea, close to, from time to time, burst into, turn one's back to, fall asleep

9	Human Communication The origin of human communication and its role in the establishment and development of civilization	Civilization and symbolism	Expository	Unreal conditions, Various conjunctions	act on, separate ~ from, similar to, stand for, warn ~ of
10	Tourist Watching in Britain Cultural differences between American and Japanese tourists from the viewpoint of a British boy who works as a tour guide	Cross-cultural understanding	Expository	Past perfect subjunctive , Participial construction of the perfect tense: <i>having</i> + past participle, <i>With</i> that expresses supplementary conditions	in the first place, on time, as a result, in advance, in turn, be curious about
11	An Observation and an Explanation An investigation into why mothers hold their babies in their right arm	Science	Expository	Nonrestrictive relative pronoun <i>which</i> , Back-shift in converting direct to indirect speech, Subject-auxiliary inversion	contrast with, insist on <i>doing</i> , carry ~ on, remind ~ of, familiar with, be liable to <i>do</i>
12	A Voice from Germany German President Weizsäcker's speech on his country's responsibility for the Second World War and contribution to keeping world peace	History and peace	Speech	Various comparative expressions, Various infinitive expressions, Various participle phrases	turn out, think of ~ as, bring about, take part in, responsible for, true to, as ~ as one can

APPENDIX D

TEXTS AND DISTRIBUTED WORKSHEETS FOR GROUPS 1 AND 2

Lesson 7 Advertising---Selling a Product, Sections 3-4, *NEW LEGEND ENGLISH COURSE II* (Suzuki, et al., 2003, pp. 73-75)

Section 3

Advertisers use many ways to get us to buy their products. One of their most successful ways is to make us feel dissatisfied with ourselves and our imperfect lives. Advertisements show us who we aren't and what we don't have. Our teeth aren't white enough. Our hair isn't shiny enough. Our clothes aren't clean enough.

Advertisements make us afraid that people won't like us if we don't use the advertised products. "Why don't I have any dates?" a pretty girl sadly asks in an advertisement. "Here," answers her friend, "try Zoom toothpaste!" Of course she tries it, and at once the whole baseball team falls in love with her. "That's nonsense," we might say. But we still buy Zoom toothpaste.

Section 4

If fear is the negative reason for buying a product, then wanting a good self-image is the positive reason for choosing it. Each of us has a mental picture of the kind of person we would like to be.

For example, a modern young woman might like to think that she looks like a beautiful movie star. A middle-aged man might want to see himself as a strong, attractive athlete. Advertisers know this. They write specific advertisements to make certain groups of people choose their product. Two people may choose different brands of toothpaste with the identical price, amount, and quality; each person believes that he or she is expressing his personality by choosing that brand.

Worksheet for Group 1

Lesson 7 Advertising---Selling a Product, Sections 3-4

10/30/06

Class _____ No. _____ Name _____

Reading Questions

1. What is one of advertisers' most successful ways to get consumers to buy their products?

.....

2. Write another example of what advertisements show we don't have, following the sentence pattern given in the text.

.....

3. What does 'That' on line 8 refer to?

.....

4. Who is your mental picture of the kind of person you would like to be?

.....

Summary Writing

Advertisers use many ways

One of their most successful ways is and

.....

Fear

Advertisements make us afraid

Wanting

Each of us has a mental picture of the kind of person we would like to be.

Advertisers to look

like their ideal person.

Worksheet for Group 2

Lesson 7 Advertising---Selling a Product, Sections 3-4

10/30/06

Class _____ No. _____ Name _____

Reading Questions

1. What is one of advertisers' most successful ways to get consumers to buy their products?

.....

2. Write another example of what advertisements show we don't have, following the sentence pattern given in the text.

.....

3. What is a positive reason for choosing a product?

.....

4. How does each person believe that he or she is expressing his or her personality?

.....

Summary Writing

Advertisers use many ways

One of their most successful ways is and

.....

Fear

Advertisements make us afraid

Wanting

Each of us has a mental picture of the kind of person we would like to be.

Advertisers to look

like their ideal person.

Lesson 9 Human Communication, Section 2, *NEW LEGEND ENGLISH COURSE II* (Suzuki, et al., 2003, pp. 97-98)

Section 2

Much of our human communication does not have to suggest meanings as visually as the cave paintings do. In fact, there may not be any clear relation between the physical characteristics of the “message” and the physical characteristics of what it is intended to mean. For example, a simple picture of an animal is visually similar to the real thing, but the word *animal* is not. When we use messages that describe things or events but that are not themselves directly similar to them, we are using symbols, the most important and most complicated characteristic of human communication.

Perhaps the picture of the Egyptian Key of Life means nothing to you. But to some people it suggests an important religious symbol. This may be because it is similar to the Christian cross, although it appeared around 1400 years before that cross. Unless we study more about Egyptian history, the Key of Life cannot be a very important symbol for us. We do not have the shared meanings necessary to understand it well.

Worksheet for Group 1

Lesson 9 Human Communication, Section 2

1/15/07

Class No. Name

Reading Questions

1. Complete the following sentence on line 5, page 98.

..., but the word animal is not

2. Give another example of symbols in your life that describe things or events but that are not themselves directly similar to them.

.....

3. What do you think is the meaning of the Egyptian symbol?

.....

4. What do we need to have with the ancient Egyptian people to understand the Key of Life well?

.....

Summary Writing

Human symbols to what they describe.

There may between a word and what it is intended to mean.

In order to communicate you have to with the person who uses that word.

Worksheet for Group 2

Lesson 9 Human Communication, Section 2

1/15/07

Class No. Name

Reading Questions

1. Complete the following sentence on line 5, page 98.

..., but the word animal is not

2. What are we using when we use messages that describe things or events but that are not themselves directly similar to them?

.....

3. Why may the Egyptian Key of Life be an important religious symbol to some people?

.....

4. What do we need to have with the ancient Egyptian people to understand the Key of Life well?

.....

Summary Writing

Human symbols to what they describe.

There may between a word and what it is intended to mean.

In order to communicate you have to with the person who uses that word.

Section 6

After a while, I saw a pattern forming. Americans would try to be friendly by not asking personal questions. They thought it too bad-mannered to ask about someone's private life. As they wanted to be polite, they would talk about themselves, thinking that I would then relax and begin talking about myself. Their reasoning was, "If I show that I trust you and feel comfortable enough with you to tell you private things about myself, then you'll relax and tell me about your private life." Their way of becoming friendly was to create an atmosphere of trust by first giving the listener private information about themselves.

The Japanese had the opposite idea. They seemed to think that the fastest way to understand people was to ask them personal questions. When they did this, they were showing their real interest in the other person, an interest which would later be returned. On the other hand, they thought that to talk about themselves was bad-mannered, for it did not show enough interest in the other person.

They were more eager to show themselves as thoughtful and sympathetic listeners than as speakers telling people private things about themselves. I often found that after I answered all their questions, then they would begin to talk openly about themselves. In other words, they were not very private. They were not asking questions of me in order to stop questions that might come at them. They were as eager to talk about themselves as Americans, but they went about it differently. To them, asking personal questions of another person was a way to be kind to that person. And I think they were also very curious about me and people of my age and status in Britain.

